

مقدمة في أساليب الاستدلال الإحصائي والتنبؤ



الأستاذ الدكتور
عادل محمود حلاوة
كلية التجارة - جامعة الإسكندرية

الأستاذة الدكتورة
إمتهال محمد حسن
كلية التجارة - جامعة الإسكندرية

الدكتورة
ليبية حسب النبي العطار
كلية التجارة - جامعة الإسكندرية



٠٠٢٠١٠ ٢٢٢٨٢٢

مقدمة فى

أساليب الإستدلال الإحصائى والتنبؤ

الأستاذة الدكتورة

إمتهال محمد حسن

كلية التجارة جامعة الإسكندرية

الأستاذ الدكتور

عادل محمود حلاوة

كلية التجارة جامعة الإسكندرية

الدكتورة

ليبية حسب النبى العطار

كلية التجارة جامعة الإسكندرية

الطبعة الأولى

2012م

الناشر

مكتبة الوفاء القالونية

محمول : 0020103738822 الإسكندرية

مقدمة

إن التطور التكنولوجي الحديث في جميع مجالات حياتنا المعاصرة من ناحية ، ودخول العالم في عصر المعلوماتية ، من ناحية أخرى ، كل هذا أدى إلى ازدياد أهمية استخدام أساليب التحليل الإحصائي في جميع مجالات المعرفة ، وعلى جميع المستويات . فعلى مستوى الاقتصاد القومي ، أو مستوى الوحدات الاقتصادية ، سواء كانت قطاع عام أو خاص ، فإن الاحتياج إلى جمع البيانات واستخراج المعلومات منها على أساس من الدراسة المنهجية الحديثة ، يعتبر من المسائل الحيوية في عصرنا الحديث . وهو ما تقوم به أساليب التحليل الإحصائي . وتستخدم هذه الأساليب في التخطيط ورسم السياسات واتخاذ القرارات والتنبؤات .

ولقد تناول هذا الكتاب " مقدمة في أساليب التحليل الإحصائي " بعرض أهم هذه الأساليب وكيفية الوصول منها إلى نتائج وتقرارات إحصائية تفيد الباحث ومتخذ القرار . وتتطلب دراسة هذه الموضوعات الإلمام بمبادئ الإحصاء الوصفي . ولقد روعي التبسيط في عرض المواضيع دون الإخلال بالمادة العلمية . ألا أن هذا الكتاب لم يتعرض لاستخدام الحاسبات الآلية في التحليل الإحصائي ، والسبب في ذلك يرجع إلى أن الكتاب مقرر على السنة الثانية بكلية التجارة ، قبل التخصص ، ونظراً للأعداد الكبيرة من الطلبة فإن الإمكانيات لا تسمح باستخدام التطبيقات على الحاسب الآلي . ألا أن هناك مقررأ في التطبيقات الإحصائية على الحاسب الآلي لطلبة التخصص في الإحصاء التطبيقي في السنة الثالثة ، حيث أعداد الطلبة صغيرة .

هذا وقد قام بتأليف هذا الكتاب كل من أ.د. امثال محمد حسن ، أ.د. عادل محمود حلاوة و د. لبيبة حسب النبي العطار.
قامت أ.د. امثال محمد حسن بكتابة الفصول الثلاثة الأولى.
قام أ.د. عادل محمود حلاوة بكتابة الفصل الرابع والفصلين السابع والثامن.
قامت د. لبيبة حسب النبي العطار بكتابة الفصلين الخامس والسادس.

هذا ويتمنى مؤلفو الكتاب لأبنائهم الطلبة حسن الأداء خلال الفصل الراسي حتى تكلل مجهوداتهم بالنجاح.

الإسكندرية في يناير ٢٠١١

المؤلفون

الفصل الأول

توزيعات المعاينة

مقدمة :

قد يكون من المفيد ، بادئ ذي بدء ، تنكسر الطالب ببعض التعريفات الهامة التي ستحتاجها في دراسته لهذا الفصل . ومن هذه التعريفات : تعريف المجتمع Population وتعريف العينة Sample .

ويمكن تعريف المجتمع بأنه " جميع " المفردات محل الدراسة سواء كانت في شكل إنسان أو حيوان أو جمد أو أشياء غير ملموسة ، وسواء كان من الممكن عدها أم لا . فيقال مثلاً : مجتمع درجات الطلبة في امتحان الإحصاء ، وهذا يعني درجات جميع الطلبة الذين دخلوا امتحان الإحصاء . أما العينة فيمكن تعريفها بأنها " مجموعة " من مفردات المجتمع . ففي مثالنا السابق إذا كان عدد الطلبة الذين تقدموا لامتحان الإحصاء ١٠٠٠ طالب ، فالمجتمع هنا هو درجات هؤلاء الطلبة الألف في مادة الإحصاء . وإذا أخذت درجات ٥٠ طالب فقط من الذين تقدموا لهذا الامتحان نقول أنه تم اختيار عينة من درجات ٥٠ طالب من مجتمع درجات الإحصاء .

وقد تكون العينة عينة عشوائية بسيطة Simple random sample حيث تكون لكل مفردة من مفردات المجتمع نفس الفرصة في تكوين العينة .

وفي أغلب الأحيان يطلق على هذا النوع " العينة العشوائية " random sample . وقد تكون العينة غير عشوائية nonrandom sample حيث لا يكون لبعض المفردات نفس الفرصة في تكوين العينة . ولإيضاح ذلك فإذا كان لدينا مجتمع مكون من ١٠٠ شخص ولدينا خمس جوائز فقط . فإذا وضعنا أسماء الأشخاص المائة في سلة وسحبنا منها خمس أسماء ، نكون إزاء عينة عشوائية . أما إذا كتبنا أسماء المائة شخص حسب الترتيب الأبجدي واختارنا الخمس أسماء الأولى ، فإن العينة تكون غير عشوائية لأن باقي الأشخاص — وهم ٩٠ شخص — لا تكون لهم فرصة في تكوين العينة .

وتسمى دراسة جميع مفردات المجتمع " بالحصص الشامل " census وهو المستخدم في التعدادات السكانية والزراعية والصناعية ... الخ . إلا أن هذا الأسلوب يتطلب تكاليف باهظة ويستهلك كثيراً من الوقت والجهد ، وقد يؤدي هذا الأسلوب إلى تلف الوحدات محل الدراسة . لذلك يلجأ كثير من الباحثين إلى استخدام أسلوب " المعاينة الإحصائية " أي أسلوب استخدام العينات .

هذا ويسمى أي مقياس إحصائي في المجتمع " بالمعلمة " parameter في حين أن أي مقياس إحصائي في العينة يسمى " إحصائية " statistic . وعادة نرسم لمعالم المجتمع بالحروف الإغريقية مثال ذلك الوسط الحسابي في المجتمع يرمز له بالرمز μ ، والانحراف المعياري في المجتمع بالرمز σ ، والنسبة في المجتمع بالرمز θ . ويجدر الإشارة هنا أن معالم المجتمع دائماً ثابتة في حين أن إحصائيات العينة فهي دائماً متغير عشوائي random variable .

وقد تكون إحصائية العينة مساوية لمعلمة المجتمع ، أو قد تكون أصغر أو أكبر منها . ويسمى الفرق بين إحصائية العينة ومعلمة المجتمع بخطأ المعاينة sampling error أو الخطأ العشوائي random error ، وهذا بافتراض أن العينة عشوائية وأنه لا يوجد أخطاء غير عشوائية . ويقصد بالأخطاء غير العشوائية الأخطاء في تجميع البيانات وتدوينها وتبويبها .

وبما أن الفرق بين إحصائية العينة ومعلمة المجتمع قد يكون سالبا أو موجبا ، لذلك يمكننا أخذ القيمة المطلقة لهذا الفرق لحساب خطأ المعاينة . أي أن :

$$\text{خطأ المعاينة} = | \text{إحصائية العينة} - \text{معلمة المجتمع} | \quad (1)$$

ويقوم هذا الفصل بدراسة توزيعات المعاينة : فيقسم إلى خمس مباحث رئيسية . يتناول المبحث الأول توزيع المجتمع ، ويتناول المبحث الثاني توزيع معاينة الوسط الحسابي \bar{x} في حالة السحب بإرجاع والسحب بدون إرجاع ، ويتناول المبحث الثالث الوسط الحسابي والانحراف المعياري لتوزيع معاينة \bar{x} ، ويتناول المبحث الرابع شكل توزيع معاينة \bar{x} ، ويتناول المبحث الخامس توزيع معاينة النسبة q .

(١ - ١) توزيع المجتمع :

بالتعريف توزيع المجتمع هو التوزيع الاحتمالي لجميع مفردات المجتمع . ويبين المثال التالي كيفية الحصول على توزيع المجتمع .

مثال (١) :

في أحد مكاتب المحاسبة يوجد ٥ موظفين ، وفيما يلي عدد سنوات الخبرة لهؤلاء الموظفين :

٦ ، ٤ ، ١٥ ، ٦ ، ١٠

والمطلوب : إيجاد التوزيع الاحتمالي لسنوات الخبرة في هذا المجتمع ومنها حساب الوسط الحسابي والانحراف المعياري لهذا المجتمع .

الحل :

من الملاحظ أن القيمة ٦ تكررت مرتين . ويبين جدول (١) التوزيع التكراري لسنوات الخبرة .

جدول (١)

التوزيع التكراري لسنوات الخبرة

سنوات الخبرة س	التكرار ك	س ك	س ^١ ك
٤	١	٤	١٦
٦	٢	١٢	٧٢
١٠	١	١٠	١٠٠
١٥	١	١٥	٢٢٥
المجموع	٥	٤١	٤١٣

ومن هذا التوزيع يمكن حساب الوسط الحسابي والانحراف المعياري كما يلي :

$$\begin{aligned} \text{الوسط الحسابي للمجتمع} : \mu &= \frac{\text{مجموع ك}}{\text{مجموع}} = \frac{41}{5} = 8,2 \text{ سنة} \\ \text{تباين المجتمع} : \sigma^2 &= \frac{1}{\text{مجموع}} [\text{مجموع ك}^2 - \frac{(\text{مجموع ك})^2}{\text{مجموع}}] \\ &= \frac{1}{5} [413 - \frac{(41)^2}{5}] = 10,36 \\ \therefore \text{الانحراف المعياري} \sigma &= \sqrt{10,36} = 3,219 \text{ سنة} \end{aligned}$$

ومن التوزيع التكراري يمكننا الحصول على التوزيع الاحتمالي
عن طريق إيجاد التكرار النسبي بقسمة التكرار على مجموع التكرارات .
ويبين الجدول التالي للتوزيع الاحتمالي لسنوات الخبرة :

جدول (٢)

التوزيع الاحتمالي لسنوات الخبرة

سنوات الخبرة س	التكرار النسبي ح (س)	س . ح (س)	س ^٢ . ح (س)
٤	$0,20 = \frac{1}{5}$	٠,٨	٣,٢
٦	$0,40 = \frac{2}{5}$	٢,٤	١٤,٤
١٠	$0,20 = \frac{1}{5}$	٢	٢٠
١٥	$0,20 = \frac{1}{5}$	٣	٤٥
· المجموع	١٠٠	٨,٢	٨٢,٦

ومن هذا التوزيع يمكن الحصول على الوسط الحسابي والانحراف
المعياري كما يلي :

الوسط الحسابي للمجتمع : $\mu = \text{م.ح.س.} (س) = ٨,٢$ سنة

تباين المجتمع : $\sigma^2 = \text{م.ح.س.}^2 (س) = \mu^2$

$$١٥,٣٦ = \mu^2 (٨,٢) - ٨٢,٦ =$$

\therefore الانحراف المعياري $\sigma = \sqrt{١٥,٣٦} = ٣,٩١٩$ سنة

ومن الواضح أن الوسط الحسابي والانحراف المعياري لهما نفس القيمة سواء كانا حسبا من التوزيع التكراري أو التوزيع الاحتمالي .

(١ - ٢) توزيع المعاينة للوسط الحسابي $\bar{س}$ في حالة السحب بإرجاع وبدون إرجاع :

رأينا في المبحث السابق كيفية الحصول على الوسط الحسابي للمجتمع μ . وكما سبق وذكرنا ، فإن الوسط الحسابي للمجتمع هو مقدار دائما ثابت ، أما الوسط الحسابي في العينة فتختلف قيمته من عينة إلى أخرى مسحوبة من نفس المجتمع ؛ ومن ثم فإن الوسط الحسابي متغيرا عشوائيا . وبما أن الوسط الحسابي في العينة متغيرا عشوائيا ، فإن له توزيع احتمالي يسمى بتوزيع المعاينة أو التوزيع العيني sampling distribution للوسط الحسابي $\bar{س}$. ونحصل على توزيع المعاينة بأخذ جميع العينات الممكنة التي لها نفس الحجم ، والتي يمكن الحصول عليها من هذا المجتمع . وفيما يلي سنتعرض لتوزيع معاينة $\bar{س}$ في حالة السحب بإرجاع وفي حالة السحب بدون إرجاع ، من خلال المثال التالي :

مثال (٢) :

بأخذ المجتمع الموجود في مثال (١) والخاص بعقد سنوات
الخبرة لموظفي أحد مكاتب المحاسبة وعددهم ٥ ، ويوضع رمز لكل
مفردة من هذه المفردات سيكون لدينا ما يلي :

سنوات الخبرة :	٦	٤	١٥	٦	١٠
الرمز :	أ	ب	ح	د	هـ

والمطلوب :

- ١ - إيجاد جميع العينات الممكنة من مفردتين التي يمكن سحبها من هذا المجتمع بإرجاع ، وحساب خطأ المعاينة ، ثم إيجاد توزيع المعاينة في هذه الحالة ومنه حساب الوسط الحسابي والانحراف المعياري .
- ٢ - إيجاد جميع العينات الممكنة للمكونة من مفردتين التي يمكن سحبها من هذا المجتمع بدون إرجاع ، ثم إيجاد توزيع المعاينة في هذه الحالة ومنه حساب الوسط الحسابي والانحراف المعياري للتوزيع .
- ١ - حالة السحب بإرجاع :

في حالة السحب بإرجاع فإن حجم المجتمع م لا يتغير عند سحب مفردة من مفردات المجتمع ، وهذا يعني أن سحب أي مفردة لا يتأثر بسحب المفردات السابقة ، ومن ثم تكون المشاهدات مستقلة عن بعضها البعض . ويمكن سحب المفردة أكثر من مرة لتكوين العينة .
فإذا أخذنا جميع العينات الممكنة المكونة من مفردتين التي يمكن سحبها من هذا المجتمع بإرجاع ، فإن :

$$\text{عدد العينات} = M^n$$

حيث : م حجم المجتمع ، ن حجم العينة .

∴ $n = 20 = 2 \times 10$ عينة .

وبين جدول (٣) هذه العينات والوسط الحسابي لكل منها .

جدول (٣)

جميع العينات الممكنة في حالة السحب بإرجاع

والوسط الحسابي لكل منها

الوسط الحسابي \bar{x}	مفردات العينة	العينة
٦	٦، ٦	أ، أ
٥	٤، ٦	أ، ب
١٠، ٥	١٥، ٦	أ، ح
٦	٦، ٦	أ، د
٨	١٠، ٦	أ، هـ
٥	٦، ٤	ب، أ
٤	٤، ٤	ب، ب
٩، ٥	١٥، ٤	ب، ح
٥	٦، ٤	ب، د
٧	١٠، ٤	ب، هـ
١٠، ٥	٦، ١٥	ح، أ
٩، ٥	٤، ١٥	ح، ب
١٥	١٥، ١٥	ح، ح
١٠، ٥	٦، ١٥	ح، د
١٢، ٥	١٠، ١٥	ح، هـ
٦	٦، ٦	د، أ
٥	٤، ٦	د، ب
١٠، ٥	١٥، ٦	د، ح
٦	٦، ٦	د، د
٨	١٠، ٦	د، هـ
٨	٦، ١٠	هـ، أ
٧	٤، ١٠	هـ، ب
١٢، ٥	١٥، ١٠	هـ، ح
٨	٦، ١٠	هـ، د
١٠	١٠، ١٠	هـ، هـ
٢٠، ٥		المجموع

ومن هذا الجدول حصلنا على الوسط الحسابي بقسمة مفردات العينة على عددها ، فمثلاً عندما كانت المفردات (٦ ، ٦) فإن :

$$\bar{س} = \frac{٦ + ٦}{٢} = ٦$$

ومن هذا الجدول يتضح لنا أن الأوساط الحسابية تختلف قيمتها من عينة إلى أخرى وفقاً لمفردات العينة ، فهو إذن متغيراً عشوائياً كما سبق وذكرنا . ومن هذا الجدول يتضح لنا أيضاً أن هذه الأوساط الحسابية تختلف عن الوسط الحسابي للمجتمع μ . ومن ثم يمكن حساب خطأ المعاينة للوسط الحسابي كما يلي :

(٢)

$$\text{خطأ المعاينة للوسط الحسابي} = | \bar{س} - \mu |$$

فمثلاً بالنسبة للعينة الأولى (أ ، أ) :

$$\text{خطأ المعاينة} = | ٨,٢ - ٦ | = ٢,٢ \text{ سنة}$$

وبالنسبة للعينة الثانية (أ ، ب) :

$$\text{خطأ المعاينة} = | ٨,٢ - ٥ | = ٣,٢ \text{ سنة}$$

وهكذا بالنسبة لباقي العينات .

وإذا أخذنا الوسط الحسابي لهذه المتوسطات نحصل على ما يلي :

$$\bar{\bar{س}} = \frac{٢٠٥}{٢٥} = ٨,٢ \text{ سنة}$$

ومن جدول (٣) يمكننا الحصول على التوزيع التكراري للأوساط الحسابية $\bar{س}$ عندما تكون العينة مكونة من مفردتين في حالة السحب بإرجاع ، كما هو موضح في جدول (٤) .

جدول (٤)

التوزيع التكراري للأوساط الحسابية (س) للعينات
عندما تكون العينة مكونة من مفرنتين في حالة السحب بإرجاع

الوسط الحسابي س	٤	٥	٦	٧	٨	٩,٥	١٠	١٠,٥	١٢,٥	١٥	المجموع
التكرار ك	١	٤	٤	٢	٤	٢٠	١	٤	٢	١	٢٥

ويقسمة التكرار على مجموع التكرارات نحصل على التكرار النسبي ، ثم نحصل على التوزيع الاحتمالي وهو في هذه الحالة توزيع المعاينة للأوساط الحسابية . كما هو مبين في جدول (٥) :

جدول (٥)

توزيع المعاينة للأوساط الحسابية (س)
عندما تكون العينة مكونة من مفرنتين في حالة السحب مع الإرجاع .

الوسط الحسابي س	٤	٥	٦	٧	٨	٩,٥	١٠	١٠,٥	١٢,٥	١٥	المجموع
التكرار النسبي ح (س)	$\frac{1}{25}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{2}{25}$	$\frac{4}{25}$	$\frac{20}{25}$	$\frac{1}{25}$	$\frac{4}{25}$	$\frac{2}{25}$	$\frac{1}{25}$	١

ولقد سبق وبيّنا في مثال (١) أنه يمكن حساب الوسط الحسابي والتباين من التوزيع التكراري أو التوزيع الاحتمالي . وفيما يلي سنستخدم التوزيع التكراري لحساب كل من الوسط الحسابي والتباين .

جدول (٦)

حساب الوسط الحسابي والانحراف المعياري

لتوزيع معاينة من في حالة السحب بإرجاع

الوسط الحسابي س	التكرار ك	س ك	س ^٢ ك
٤	٩	٤	١٦
٥	٤	٢٠	١٠٠
٦	٤	٢٤	١٤٤
٧	٢	١٤	٩٨
٨	٤	٣٢	٢٥٦
٩,٥	٢	١٩	١٨٠,٥
١٠	١	١٠	١٠٠
١٠,٥	٤	٤٢	٤٤١
١٢,٥	٢	٢٥	٣١٢,٥
١٥	١	١٥	٢٢٥
المجموع	٢٥	٢٠٥	١٨٧٣

ومن جدول (٦) يمكننا الحصول على الوسط الحسابي لتوزيع

المعاينة ، والذي يرمز له بالرمز (μ) ، كما يلي :

$$\mu = \frac{\sum س ك}{\sum ك} = \frac{٢٠٥}{٢٥} = ٨,٢ \text{ مئة}$$

ولقد حصلنا على نفس هذه القيمة بحساب الوسط الحسابي للأوساط

الحسابية من جدول (٣) ، حيث : $\bar{س} = ٨,٢ \text{ مئة}$.

وفي الواقع فإن الوسط الحسابي للتوزيع العيني \bar{x} ما هو إلا الوسط الحسابي للأوساط الحسابية ، أي أن :

(٣)

$$\bar{x} = \bar{\bar{x}}$$

وفي مثال (١) وجدنا أن الوسط الحسابي للمجتمع $\mu = ٨,٢$ سنة أيضاً . وهذا التساوي بين الوسط الحسابي للتوزيع العيني (\bar{x}) والوسط الحسابي للمجتمع (μ) ليس وليد للصدفة ولكن هذا التساوي هو نتيجة لخاصية مهمة ألا وهي :

الوسط الحسابي للأوساط الحسابية للعينات يساوي الوسط الحسابي للمجتمع أي أنه

(٤)

$$\mu = \bar{\mu}$$

ومن جدول (٦) يمكننا الحصول على تباين توزيع معاينة \bar{s} ، والذي يرمز له بالرمز (σ^2) كما يلي :

$$\sigma^2 = \frac{1}{n} \left[\sum_{k=1}^n \frac{(x_k - \bar{x})^2}{f_k} \right] = \frac{1}{25} \left[\frac{(205)^2}{25} - 1,873 \right] = 7,68$$

$$\therefore \sigma = \sqrt{7,68} = 2,771 \text{ سنة}$$

ويسمى الانحراف المعياري لتوزيع المعاينة بالخطأ المعياري . standard error

ومن الملاحظ أن الانحراف المعياري للتوزيع العيني
($\sigma = ٢,٧٧١$ سنة) لا يساوي الانحراف المعياري للمجتمع الذي سبق
وحصلنا على قيمته في مثال (١) حيث ($\sigma = ٣,٩١٩$ سنة) . ولكن
توجد علاقة بينهما ، ألا وهي :

$$(٥) \quad \boxed{\frac{\sigma}{\sqrt{n}} = \sigma_{\bar{x}}}$$

$$\text{وبالفعل : } \sigma_{\bar{x}} = \frac{٣,٩١٩}{\sqrt{٢}} = ٢,٧٧١ \text{ سنة}$$

ولقد سبق وحصلنا على هذه القيمة عند حساب $\sigma_{\bar{x}}$ من جدول (٦) .
وتكون هذه العلاقة بين الانحراف المعياري لتوزيع معاينة \bar{x} وبين
الانحراف المعياري للمجتمع - علاقة (٥) - صحيحة إذا تم سحب
العينات بإرجاع من مجتمع محدود الحجم *finite population* ، حيث لا
يؤثر سحب مفردة من المجتمع على سحب مفردة أخرى ، ومن ثم يكون
هذين الحدثين مستقلين إحصائياً ، وتكون العلاقة (٥) صحيحة أيضاً في
حالة المجتمعات اللانهائية الحجم - سواء كان السحب بإرجاع و بدون
إرجاع - حيث لا يتأثر حجم المجتمع بسحب مفردة منه ، ومن ثم يكون
حدث سحب مفردة مستقلاً عن حدث سحب مفردة أخرى . ويتحقق هذا
الاستقلال الإحصائي عندما يكون حجم المجتمع n كبيراً بالنسبة لحجم
العينة n ، أي عندما $\frac{n}{N} > ٠,٠٥$ م^(١).

(1) Kohler, H., "Statistics for Business & Economics", Harper Collins College
Publisher, 1994, pp. 305, 306.

Mann P. S., "Statistics for Business & Economics", John Wiley & Sons.,
Inc., 1995, p. 371.

و خلاصة القول :

عندما يكون السحب بإرجاع ، أو عندما تكون النسبة بين حجم العينة n وحجم المجتمع M هي : $\frac{n}{M} > 0.05$ ، أي في حالة تحقق الاستقلال الإحصائي ، فإن :

$$(6) \quad \frac{\sigma}{\sqrt{n}} = \sigma$$

وفي أغلب الأحيان يكون حجم العينة صغيرا نسبيا بالنسبة لحجم المجتمع مما يجعل العلاقة (٦) أكثر استخداما في الحياة العملية .

٢ - حالة السحب بدون إرجاع :

في حالة السحب بدون إرجاع فإن أي مفردة في المجتمع لا يمكن أن تسحب إلا مرة واحدة لتكوين العينة . فإذا أخذنا جميع العينات الممكنة للمكونة من مفردتين بدون إرجاع فسيكون لدينا ما يلي :

أ ب	أ د	أ هـ	ب د	ب هـ	ج د	ج هـ	د هـ
٤٠٦	١٥٠٦	٦٠٦	١٠٠٦	١٥٠٤	٦٠٤	١٠٠٤	٦٠١٥

أي أن لدينا ١٠ عينات . ويمكن الحصول على عدد العينات الكلية الممكنة بدون إرجاع عن طريق التوافيق :

$$M(M-1)(M-2) \dots (M-n+1)$$

عدد العينات الكلية الممكنة = ${}^M C_n = \frac{M(M-1)(M-2) \dots (M-n+1)}{1 \times 2 \times 3 \times \dots \times n}$

$$= {}^{10} C_2 = \frac{10 \times 9}{1 \times 2} = 45 \text{ عينات .}$$

وبين جدول (٧) جميع العينات الممكنة المكونة من مفردتين
والوسط الحسابي لكل منها :

جدول (٧)

جميع العينات الممكنة المكونة من مفردتين في حالة السحب بدون إرجاع
والوسط الحسابي لكل منها

الوسط الحسابي \bar{x}	مفردات العينة	العينة
٥	٤ ، ٦	أ ب
١٠,٥	١٥ ، ٦	أ ح
٦	٦ ، ٦	أ د
٨	١٠ ، ٦	أ هـ
٩,٥	١٥ ، ٤	ب ح
٥	٦ ، ٤	ب د
٧	١٠ ، ٤	ب هـ
١٠,٥	٦ ، ١٥	ج د
١٢,٥	١٠ ، ١٥	ج هـ
٨	١٠ ، ٦	د هـ
٨٢		المجموع

وحصلنا على الوسط الحسابي بقسمة مفردات العينة على عددها ، فمثلاً
عندما كانت المفردات (٤ ، ٦) فإن :

$$\bar{x} = \frac{4 + 6}{2} = 5$$

ومن جدول (٧) يتضح لنا أن الأوساط الحسابية تختلف قيمتها من عينة إلى أخرى وفقاً لمفردات العينة . كما أن كل من هذه الأوساط الحسابية يختلف عن الوسط الحسابي للمجتمع μ ، حيث $\mu = ٨,٢$ سنة كما سبق وحصلنا عليها في مبحث (١ - ١) . ومن جدول (٧) يمكننا الحصول على التوزيع التكراري للأوساط الحسابية كما هو موضح فسي جدول (٨) .

جدول (٨) .

التوزيع التكراري للأوساط الحسابية \bar{x}

عندما تكون العينة مكونة من مفردتين ، في حالة السحب بدون إرجاع

التكرار ك	الوسط الحسابي \bar{x}
٢	٥
١	٦
١	٧
٢	٨
١	٩,٥
٢	١٠,٥
١	١٢,٥
١٠	المجموع

وبقسمة التكرار على مجموع التكرارات نحصل على التكرار النسبي ، وهو يمثل الاحتمال في كل فئة . ويبين جدول (٩) توزيع المعاينة للأوساط الحسابية \bar{x} عندما تكون العينة مكونة من مفردتين .

جدول (٩)

توزيع المعاينة للأوساط الحسابية \bar{S}
عندما تكون العينة مكونة من مفرنتين

التكرار النسبي ح (\bar{S})	الوسط الحسابي \bar{S}
$\frac{2}{10}$	٥
$\frac{1}{10}$	٦
$\frac{1}{10}$	٧
$\frac{2}{10}$	٨
$\frac{1}{10}$	٩,٥
$\frac{2}{10}$	١٠,٥
$\frac{1}{10}$	١٢,٥
١,٠٠	المجموع

وكما سبق وذكرنا ، فإنه يرمز للوسط الحسابي لتوزيع معاينة \bar{S}
بالرمز $\mu_{\bar{S}}$ أو \bar{S} . ويرمز للانحراف المعياري لتوزيع معاينة \bar{S}
بالرمز $\sigma_{\bar{S}}$.

يمكن الحصول على \bar{S} و $\sigma_{\bar{S}}$ باستخدام جدول (٧) ، أي
حالة القيم غير المبوبة ، كما هو موضح من جدول (١٠) .

جدول (١٠)

حساب الوسط الحسابي والانحراف المعياري
للأوساط الحسابية في حالة السحب بدون إرجاع

الوسط الحسابي \bar{X}	$\bar{X} - \bar{X}$	$(\bar{X} - \bar{X})^2$
٥	٣,٢ -	١٠,٢٤
١٠,٥	٢,٣	٥,٢٩
٦	٢,٢ -	٤,٨٤
٨	٠,٢ -	٠,٠٤
٩,٥	١,٣	١,٩٩
٥	٣,٢ -	١٠,٢٤
٧	١,٢ -	١,٤٤
١٠,٥	٢,٣	٥,٢٩
١٢,٥	٤,٣	١٨,٤٩
٨	٠,٢ -	٠,٠٤
٨٢	صفر	٥٧,٦٠

$$\bar{X} = \frac{\sum \bar{X}}{\text{عدد العينات}} = \frac{٨٢}{١٠} = ٨,٢ \text{ سنة}$$

$$\sigma^2 = \frac{\sum (\bar{X} - \bar{X})^2}{\text{عدد العينات}} = \frac{٥٧,٦٠}{١٠} = ٥,٧٦$$

$$\sigma = \sqrt{٥,٧٦} = ٢,٤ \text{ سنة}$$

كما يمكن الحصول على \bar{S} و σ باستخدام جدول (٨) ،
أي حالة القيم المبوبة ، كما هو موضح من جدول (١١) .

جدول (١١)

حساب الوسط الحسابي والانحراف المعياري
لتوزيع معاينة \bar{S} في حالة السحب بدون إرجاع

الوسط الحسابي \bar{S}	التكرار ك	S ك	S^2 ك
٥	٢	١٠	٥٠
٦	١	٦	٣٦
٧	١	٧	٤٩
٨	٢	١٦	١٢٨
٩,٥	١	٩,٥	٩٠,٢٥
١٠,٥	٣	٢١	٢٢٠,٥٠
١٤,٥	١	١٠٠	١٥٦,٢٥
المجموع	١٠	٨٢	٧٣٠

$$\mu = \frac{\sum S \cdot K}{\sum K} = \frac{82}{10} = 8,2 \text{ سنة}$$

وكما سبق أن وجدنا في حالة السحب مع إرجاع ، فإن الوسط الحسابي
للتوزيع العيني في حالة السحب بدون إرجاع يساوي للوسط الحسابي
للمجتمع . أي أن :

(٧)

$$\mu = \bar{\mu}$$

أما تبين توزيع معاينة \bar{m} في حالة السحب بدون إرجاع ،

$$\sigma_{\bar{m}}^2 = \frac{1}{m} \left[\frac{m(m-1)}{m} - \frac{1}{m} \right]$$

$$0,76 = \left[\frac{1}{10} - \frac{1}{10} \right] \frac{1}{m}$$

$$\sigma_{\bar{m}} = \sqrt{0,76} = 0,87$$

وبوجه عام ، عندما يتم السحب بدون إرجاع ، فإن سحب مفردة من المجتمع لتكوين العينة يؤثر على سحب أي مفردة أخرى وهنا ، يكون الحدثين غير مستقلين . ويتحقق عدم الاستقلال الإحصائي أيضا عندما يكون حجم المجتمع صغير نسبيا بالنسبة لحجم العينة أي عندما يتحقق الشرط : $\frac{n}{N} \leq 0,05$ ^(١) .

عندما يتم السحب بدون إرجاع ، أو عندما تكون النسبة بين حجم العينة n وحجم المجتمع N هي : $\frac{n}{N} \leq 0,05$ ، أي في حالة عدم تحقق الاستقلال الإحصائي ، فإن :

$$\sigma_{\bar{m}} = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N-n}{N-1}} \quad (٨)$$

ويسمى المقدار $\frac{N-n}{N-1}$ بمعامل التصحيح correction factor ولقد

سبق ووجدنا من جدول (١) ، (٢) أن : $\sigma = 3,919$

$$\sigma_{\bar{m}} = \frac{3,919}{\sqrt{2}} \sqrt{\frac{2-0}{2-1}} = 2,77 \approx 2,8 \text{ سنة .}$$

^(١) Kohler,H., op.cit. Pp. 305, 306

مثال (٣) :

إذا كان حجم المجتمع ٦٠٠٠ مفردة ، وكان الوسط الحسابي والانحراف المعياري في هذا المجتمع هما : ٢٦ ، ٣ على التوالي ، فالمطلوب حساب كل من الوسط الحسابي والانحراف المعياري لتوزيع معاينة \bar{S} ، إذا كان حجم العينة :

أولاً : ٣٠ مفردة

ثانياً : ٥٠٠ مفردة .

الحل :

أولاً : $N = 30$ ، $M = 6000$ ، $\mu = 26$ ، $\sigma = 3$

$\mu_{\bar{S}} = 26$ ، $\sigma_{\bar{S}} = 3$

الوسط الحسابي لتوزيع معاينة \bar{S} :

$$\mu_{\bar{S}} = \mu = 26$$

في هذه الحالة $\frac{N}{M} = \frac{30}{6000} = 0,005 > 0,05$

ومن ثم فإن الانحراف المعياري لتوزيع معاينة \bar{S} نحصل عليه من

القانون :

$$\sigma_{\bar{S}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{30}} = 0,548$$

ثانياً : $\bar{N} = 500$ ، $\bar{M} = 6000$ ، $\bar{\mu} = 26$ ، $\bar{\sigma} = 3$

$$\bar{\mu} = \bar{\mu} = 26$$

الوسط الحسابي لتوزيع معاينة \bar{S} :

$$\bar{\mu} = \bar{\mu} = 26$$

$$\text{في هذه الحالة } \frac{\bar{N}}{\bar{M}} = \frac{500}{6000} = 0,083 < 0,05$$

ومن ثم فإن الانحراف المعياري لتوزيع معاينة \bar{S} نحصل عليه من

القانون :

$$\bar{\sigma} = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N-M}{1-M}}$$

$$= \frac{3}{500\sqrt{1-6000}}$$

$$= \frac{3}{500\sqrt{0,9575}} = 0,128$$

(١ - ٣) الوسط الحسابي والانحراف المعياري لتوزيع

معاينة \bar{S} :

لقد عرفنا في بداية هذا الفصل العينة العشوائية البسيطة حيث

تكون لكل مفردة نفس الفرصة في تكوين العينة . ومن ثم فإن كل مفردة

من مفردات العينة تكون مستقلة عن باقي المفردات .

نظرية ١ :

العينة العشوائية البسيطة هي تلك العينة التي تكون كل مشاهداتها (س_١ ، س_٢ ، . . . ، س_ن) مستقلة . ويكون توزيع كل مشاهدة س هو نفسه توزيع المجتمع الذي سحبت منه ، أي أن :

$$\text{توزيع س}_1 = \text{توزيع س}_2 = \dots = \text{توزيع س}_n = \text{توزيع المجتمع}$$

ومن ثم فإن كل مشاهدة في العينة يكون :

$$\text{وسطها الحسابي} = \text{الوسط الحسابي للمجتمع الذي سحبت منه}$$

$$\text{وانحرافها المعياري} = \text{الانحراف المعياري للمجتمع الذي سحبت منه}$$

وبدراسة المتغيرات العشوائية فإن :

$$(٩) \quad (س + ص) = ت (س) + ت (ص)$$

$$(١٠) \quad ت (أس + ب ص) = أ ت (س) + ب ت (ص)$$

وإذا كانت س ، ص مستقلتان فإن :

$$(١١) \quad \text{تباين} (س + ص) = \text{تباين س} + \text{تباين ص}$$

$$(١٢) \quad \text{تباين} (أس + ب ص) = أ^2 \text{تباين س} + ب^2 \text{تباين ص}$$

ولقد بينا في المبحث السابق أن الوسط الحسابي لتوزيع معاينة $\bar{س}$ يساوي الوسط الحسابي للمجتمع ، وفيما يلي برهان هذه المعادلة .

إذا كان لدينا المشاهدات الآتية :

س_١ ، س_٢ ، . . . ، س_ن

فالتعريف ، الوسط الحسابي لهذه القيم هو :

$$\bar{s} = \frac{1}{n} (s_1 + s_2 + \dots + s_n)$$

$$= \frac{1}{n} (s_1) + \frac{1}{n} (s_2) + \dots + \frac{1}{n} (s_n)$$

فإذا كانت كل من : s_1, s_2, \dots, s_n متغيرات عشوائية فإن \bar{s} بدوره يكون متغيراً عشوائياً . وباستخدام معادلة (١٠) ، فإن

$$ت (\bar{s}) = \frac{1}{n} ت (s_1) + \frac{1}{n} ت (s_2) + \dots + \frac{1}{n} ت (s_n)$$

$$\therefore ت (\bar{s}) = \frac{1}{n} [ت (s_1) + ت (s_2) + \dots + ت (s_n)] \quad (١٣)$$

وباستخدام نظرية ١ ، فإن كل مشاهدة (س) لها نفس توزيع المجتمع الذي سحبت منه ووسطها الحسابي يساوي الوسط الحسابي للمجتمع μ . أي أن :

$$ت (s_1) = ت (s_2) = \dots = ت (s_n) = \mu \quad (١٤)$$

ومن ثم ، فإن معادلة (١٣) تصبح :

$$ت (\bar{s}) = \frac{1}{n} \underbrace{[\mu + \mu + \dots + \mu]}_{n \text{ مرة}}$$

$$= \frac{1}{n} (\mu \cdot n) = \mu$$

$$(١٥) \quad \boxed{ت (\bar{s}) = \mu} \quad \text{أي أن :}$$

وهذا معناه أن الوسط الحسابي لتوزيع معاينة \bar{s} يساوي الوسط الحسابي للمجتمع . وبما أن الوسط الحسابي \bar{s} متغيراً عشوائياً قيمته تختلف من عينة إلى أخرى ، فتارة تزيد عن قيمة μ وتارة تنخفض عن قيمة μ ، مما يدفعنا إلى معرفة تباين توزيع معاينة \bar{s} .

ولقد سبق وبيننا أن :

$$\bar{s} = \frac{1}{n} (s_1 + s_2 + \dots + s_n)$$

أي أن :

$$\bar{s} = \frac{1}{n} (s_1) + \frac{1}{n} (s_2) + \dots + \frac{1}{n} (s_n)$$

وبما أن s_1 ، s_2 ، ... ، s_n متغيرات عشوائية مستقلة ، وباستخدام معادلة (١١) :

$$\bar{s} = \text{تباين } \left(\frac{1}{n} s_1 \right) + \text{تباين } \left(\frac{1}{n} s_2 \right) + \dots + \text{تباين } \left(\frac{1}{n} s_n \right)$$

وبما أن s_1 ، s_2 ، ... ، s_n مستقلة ، وباستخدام قانون (١٢) ، فإن :

$$\bar{s} = \left(\frac{1}{n} \right)^2 \text{تباين } s_1 + \left(\frac{1}{n} \right)^2 \text{تباين } s_2 + \dots + \left(\frac{1}{n} \right)^2 \text{تباين } s_n \quad (١٦)$$

ووفقا لنظرية ١ ، فإن الانحراف المعياري لكل مشاهدة يساوي الانحراف المعياري للمجتمع الذي سحبت منه ، أي أن :

$$\text{انحراف معياري } s_1 = \text{انحراف معياري } s_2 = \dots = \text{انحراف معياري } s_n$$

$$\sigma =$$

وهذا يعني أن :

$$\text{تباين } s_1 = \text{تباين } s_2 = \dots = \text{تباين } s_n = \sigma^2 \quad (١٧)$$

وباستخدام معادلة (١٧) ، تصبح معادلة (١٦) كما يلي :

$$\bar{s} = \sigma^2 \left(\frac{1}{N} \right) + \dots + \sigma^2 \left(\frac{1}{N} \right) + \sigma^2 \left(\frac{1}{N} \right)$$

$$= \underbrace{(\sigma^2 + \dots + \sigma^2 + \sigma^2)}_{\text{ن مرة}} \frac{1}{N}$$

$$\text{أي أن : } \bar{s} = \frac{1}{N} (\sigma^2 \text{ ن})$$

$$\bar{s} = \frac{\sigma^2}{N} \quad \text{تباين } \bar{s}$$

(١٨)

$$\boxed{\frac{\sigma^2}{N} = \bar{s}}$$

وبأخذ الجذر التربيعي للطرفين ، فإن الانحراف المعياري لتوزيع معاينة \bar{s} (أي الخطأ المعياري) هو :

(١٩)

$$\boxed{\frac{\sigma}{\sqrt{N}} = \bar{s}}$$

(١ - ٤) شكل توزيع معاينة \bar{s} :

من نظرية ١ ، رأينا أن توزيع كل مشاهدة s ، في العينة العشوائية البسيطة ، لها نفس توزيع المجتمع الذي سحبت منه . والسؤال الذي يثار هنا ما هو شكل توزيع معاينة \bar{s} . وهنا يجب التفرقة بين حالتين : حالة العينات المسحوبة من مجتمع يتبع التوزيع المعتدل وحالة العينات المسحوبة من مجتمع لا يتبع التوزيع المعتدل .

١ - العينات المسحوبة من مجتمع يتبع التوزيع المعتدل :

نظرية ٢ :

إذا كانت : s_1, s_2, \dots, s_n من مفردات عينة عشوائية مكونة من n مفردة ، ومسحوبة من مجتمع معتدل (μ, σ^2) ، فإن توزيع معاينة \bar{s} يكون أيضا معتدلا $(\mu, \frac{\sigma^2}{n})$.

وهذا يعني إذا المجتمع معتدل ومتوسطه μ وتباينه σ^2 ، فإن توزيع معاينة الوسط الحسابي لجميع العينات الممكنة التي لها نفس الحجم ، والمسحوبة من هذا المجتمع ، يكون له الخصائص التالية :

١ - هذا التوزيع معتدل .

٢ - الوسط الحسابي لهذا التوزيع يساوي الوسط الحسابي للمجتمع

$$\text{أي : } \mu = \bar{\mu}$$

٣ - تباين هذا التوزيع يساوي تباين المجتمع مقسوما على حجم العينة

$$\text{أي : } \frac{\sigma^2}{n} = \bar{\sigma}^2$$

وبأخذ الجذر التربيعي للطرفين ، فإن الانحراف المعياري لتوزيع المعاينة (المسمى بالخطأ المعياري) يساوي الانحراف المعياري للمجتمع مقسوما على الجذر التربيعي لحجم العينة :

$$\frac{\sigma}{\sqrt{n}} = \bar{\sigma}$$

ومن دراستنا للتوزيع المعتدل ع (μ ، σ^2) ، رأينا أنه يمكن حساب الاحتمالات المختلفة عن طريق إيجاد المساحات المناظرة تحت المنحنى المعتدل المعياري ع (صفر ، ١) ، أي المنحنى المعتدل الذي يكون وسطه الحسابي صفر وانحرافه المعياري ١ . لذلك نحول قيم المتغير العشوائي س إلى قيم معيارية Z ، حيث :

$$\frac{\mu - \bar{S}}{\sigma} = Z$$

ثم نستخدم جداول المنحنى المعتدل المعياري لإيجاد المساحات المطلوبة .

إذا كان لدينا عينة عشوائية مكونة من ن مفردة ، وكان توزيع معاينة وسطها الحسابي معتدلاً ع (μ ، $\frac{\sigma}{\sqrt{n}}$) ، وكان المطلوب إيجاد الاحتمالات المختلفة تحت المنحنى المعتدل ، فإننا نقوم بتحويل قيم س إلى قيم معيارية Z حيث :

$$\frac{\mu - \bar{S}}{\frac{\sigma}{\sqrt{n}}} = Z$$

ثم نستخدم جداول المنحنى المعتدل المعياري لإيجاد المساحات المطلوبة .

مثال (٤) :

في أحد امتحانات مادة الإحصاء في السنة الثانية لكلية التجارة ، كانت درجات الطلبة تتوزع توزيعاً معتدلاً ، وسطه الحسابي ٧٠ درجة وانحرافه المعياري ١٠ درجات .

أولاً : إذا سحبت ورقة امتحان واحدة عشوائيا ، أوجد احتمال أن يكون هذا الطالب حاصل على درجة أكبر من ٧٥ .

ثانيا : إذا سحبت عينة عشوائية من ٩ طلاب ، احسب الوسط الحسابي والخطأ المعياري لتوزيع معاينة \bar{x} ، ثم أوجد احتمال أن يكون الوسط الحسابي لهذه العينة أكبر من ٧٥ درجة .

ثالثا : إذا سحبت عينة عشوائية من ٤٩ طالب ، احسب للوسط الحسابي والخطأ المعياري لتوزيع معاينة \bar{x} ، ثم أوجد احتمال أن يكون الوسط الحسابي لهذه العينة ينحصر بين ٦٨ درجة و ٧٢ درجة .

الحل :

$$\mu = 70 \text{ درجة} , \sigma = 10 \text{ درجات}$$

أولاً : بما أن درجات الطلبة تتوزع توزيعاً معتملاً ، فإن درجة الطالب المسحوبة من هذا المجتمع تتوزع هي أيضاً توزيعاً معتملاً .

$$P(x < 70) = P\left(\frac{x - \mu}{\sigma} < \frac{70 - 70}{10}\right)$$

$$= P(Z < 0)$$

$$= 1 - \Phi(0)$$

$$= 1 - 0.5$$

$$= 0.5$$

ثانيا : $n = 9$ ، $\mu = 70$ درجة ، $\sigma = 10$ درجات .

الوسط الحسابي لتوزيع معاينة \bar{x} :

$$\mu = \bar{x} = 70 \text{ درجة}$$

الخطأ المعياري لتوزيع معاينة \bar{x} :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{9\sqrt{}} = 3,33 \text{ درجة}$$

$$P(\bar{x} < 70) = P\left(Z < \frac{70 - 70}{3,33}\right)$$

$$= P(Z < 0)$$

$$= 1 - \Phi(0)$$

$$= 1 - 0,9332$$

$$= 0,0668$$

ثالثاً : $n = 49$ ، $\mu = 70$ درجة ، $\sigma = 10$ درجات .

الوسط الحسابي لتوزيع معاينة \bar{x} :

$$\mu = \bar{x} = 70 \text{ درجة}$$

الخطأ المعياري لتوزيع معاينة \bar{x} :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{7\sqrt{}} = 1,43 \text{ درجة}$$

$$P(70 \leq \bar{x} \leq 72) = P\left(\frac{70 - 70}{1,43} \leq Z \leq \frac{72 - 70}{1,43}\right)$$

$$= P(-1,4 \leq Z \leq 1,4)$$

$$= \Phi(1,4) - \Phi(-1,4)$$

$$= [\Phi(1,4) - 1] - \Phi(-1,4)$$

$$= 1 - \Phi(1,4) - \Phi(-1,4)$$

$$= 1 - 0,9192$$

$$= 0,0808$$

٢ - العينات المسحوبة من مجتمع لا يتبع التوزيع المعتدل :

قد يحدث في كثير من الأحيان أن يكون المجتمع الذي تسحب منه العينات ليس معتدلاً . فقد يكون ملتوياً نحو اليمين أو نحو اليسار . في هذه الحالة نطبق نظرية النهاية المركزية Central limit theorem .

نظرية ٣ :

نظرية النهاية المركزية :

إذا كانت : s_1, s_2, \dots, s_n مفردات عينة عشوائية مكونة من n مفردة ، ومسحوبة من مجتمع لا يتبع التوزيع المعتدل ، ووسطه الحسابي μ وتباينه σ^2 ، فإن توزيع معاينة \bar{s} يقترب من التوزيع المعتدل $E(\mu, \frac{\sigma}{\sqrt{n}})$ كلما ازداد حجم العينة (n) .
تتطبق نظرية النهاية المركزية على المجتمعات الكبيرة فقط ، أي عندما $n \leq 30$.

وفقاً لنظرية النهاية المركزية ، فإذا كان توزيع المجتمع الذي تسحب فيه العينات ليس معتدلاً ، فإن شكل توزيع معاينة \bar{s} لا يكون معتدلاً بالضبط ، ولكنه يكون قريباً من الاعتدال عندما يكون حجم العينة كبيراً ($n \leq 30$) .

ومن الملاحظ أن : $\mu_s = \mu$

وأن : $\frac{\sigma}{\sqrt{n}} = \frac{\sigma_s}{\sqrt{n}}$

أي : $\frac{\sigma}{\sqrt{n}} = \frac{\sigma_s}{\sqrt{n}}$

ولكن يجب أن نتذكر أنه لكي تكون : $\sigma = \frac{\sigma}{\sqrt{n}}$ يجب تحقيق الشرط القائل بأن نسبة حجم العينة إلى حجم المجتمع : $\frac{n}{m} > 0,05$ ، أما إذا كانت $\frac{n}{m} \leq 0,05$ فيجب استخدام معامل التصحيح عند حساب σ كما سبق وذكرنا في معادلة (٨) .

مثال (٥) :

في أحد المصانع كان الوسط الحسابي لأجور العمال ٣٠٠ جنيهه وانحرافه المعياري ٥٠ جنيهه . ولقد سحبت عينة عشوائية من ٤٠ عامل من عمال هذا المصنع ، أوجد :

أولاً : احتمال أن يكون الوسط الحسابي للعينة أقل من ٢٧٥ جنيهه .

ثانياً : احتمال أن يكون الوسط الحسابي للعينة أكبر من ٣٢٠ جنيهه .

ثالثاً : احتمال أن يكون الوسط الحسابي للعينة بين ٢٨٠ و ٣١٥ جنيهه .

الحل :

هنا $\mu = 300$ جنيهه ، $\sigma = 50$ جنيهه ، $n = 40$.

وحيث أن توزيع مجتمع أجور العمال غير معطى ، أي غير معلوم ، ونظراً لأن حجم العينة كبير ($n > 30$) ، إذاً يمكننا تطبيق نظرية النهاية المركزية . ومن ثم فإن توزيع معاينة \bar{x} يكون معتدلاً تقريباً بوسط حسابي : $\mu = \mu = 300$ جنيهه

$$\text{وخطأ معياري : } \sigma = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{40}} = 7,91 \text{ جنيهه}$$

$$\text{أولاً : } \mathcal{C} = (\overline{S} > 270) \mathcal{C} = \left(\frac{300 - 270}{\sqrt{91}} > \frac{\mu - \overline{S}}{\overline{S}\sigma} \right) \mathcal{C}$$

$$\mathcal{C} = (2.16 > Z) \mathcal{C} =$$

$$= 1 - \Phi(2.16)$$

$$= 1 - 0.992$$

$$= 0.008$$

$$\text{ثانياً : } \mathcal{C} = (\overline{S} < 220) \mathcal{C} = \left(\frac{300 - 220}{\sqrt{91}} < \frac{\mu - \overline{S}}{\overline{S}\sigma} \right) \mathcal{C}$$

$$\mathcal{C} = (2.03 < Z) \mathcal{C} =$$

$$= 1 - \Phi(2.03)$$

$$= 1 - 0.9943$$

$$= 0.0057$$

ثالثاً :

$$\mathcal{C} = (280 \leq \overline{S} \leq 310) \mathcal{C} = \left(\frac{300 - 310}{\sqrt{91}} \geq Z \geq \frac{300 - 280}{\sqrt{91}} \right) \mathcal{C}$$

$$\mathcal{C} = (1.90 \geq Z \geq 2.03) \mathcal{C} =$$

$$= \Phi(2.03) - \Phi(1.90)$$

$$= \Phi(2.03) - \Phi(1.90)$$

$$= 0.9731 - (1 - 0.9943)$$

$$= 0.9713 + 0.9943 - 1$$

$$= 0.9656$$

(١ - ٥) توزيع معاينة النسبة ق :

في كثير من الأحيان يتم تقسيم المجتمع إلى نوعين من المفردات :
المفردات التي تتصف بصفة معينة والمفردات التي لا تتصف بهذه الصفة .
فمثلاً قد ينقسم المجتمع إلى مدخن وغير مدخن ، أو إلى إناث وذكور ،
وقد ينقسم مجتمع إنتاج مصنع معين إلى إنتاج معيب وإنتاج غير معيب .
فإذا سحبت مفردة من أحد هذه المجتمعات بطريقة عشوائية ، فإن :

احتمال تمتع هذه المفردة بصفة معينة = نسبة هذه الظاهرة في المجتمع = θ

أما احتمال عدم تمتع المفردة المسحوبة بهذه الصفة = $1 - \theta$.
فمثلاً إذا كانت نسبة المدخنين في مجتمع معين هي $\theta = 0,4$ ، فإن نسبة
غير المدخنين في هذا المجتمع تساوي $(1 - \theta) = 0,4 - 1 = 0,6$

وإذا سحبنا عينة عشوائية حجمها n من هذا المجتمع ، وإذا كان
 s هو المتغير العشوائي الذي يمثل عدد الذين يتصفون بهذه الصفة في
المجتمع ، فإن s تكون متغيراً عشوائياً له توزيع ذي الحدين . ويمكن
حساب النسبة (ق) في العينة بقسمة عدد المفردات التي تتمتع بالصفة
العينة على حجم العينة ، أي أن :

$$ق = \frac{\text{عدد المفردات التي تتمتع بالصفة}}{\text{حجم العينة}} = \frac{s}{n} \quad (٢٠)$$

وبما أن عدد المفردات التي تتمتع بالصفة يختلف من عينة إلى أخرى، إذن
فالنسبة ق هي متغير عشوائي له توزيع احتمالي هو توزيع معاينة نسبة
العينة ق . ويوضح مثال (٦) هذا للتوزيع .

مثال (٦) :

في أحد فصول الدراسات العليا كان عدد الطلبة المتقدمين لامتحان الإحصاء ٥ طلبة . وكانت نتيجة الامتحان كما يلي :

رقم جلوس الطالب	١	٢	٣	٤	٥
النتيجة	ناجح	راسب	ناجح	ناجح	راسب

ويمكننا حساب نسبة النجاح في هذا المجتمع كما يلي :

$$\theta = \frac{3}{5} = 0,60$$

وإذا أخذنا جميع العينات الممكنة المكونة من مفردتين ، والتي يمكن سحبها

من هذا المجتمع بدون إرجاع ، فإن :

$$\text{عدد العينات الكلية الممكنة} = \text{ق}^2 = \frac{4 \times 5}{1 \times 2} = 10 \text{ عينات}$$

ويبين جدول (١٢) جميع العينات الممكنة في هذه الحالة ، والنسبة في كل منها .

جدول (١٢)

جميع العينات الممكنة ونسبة كل منها

نسبة العينة ق	مفردات العينة	العينة
٠,٥	ناجح ، راسب	٢ ، ١
١	ناجح ، ناجح	٣ ، ١
١	ناجح ، ناجح	٤ ، ١
٠,٥	ناجح ، راسب	٥ ، ١
٠,٥	راسب ، ناجح	٣ ، ٢
٠,٥	راسب ، ناجح	٤ ، ٢
صفر	راسب ، راسب	٥ ، ٢
١	ناجح ، ناجح	٤ ، ٣
٠,٥	ناجح ، راسب	٥ ، ٣
٠,٥	ناجح ، راسب	٥ ، ٤

من هذا الجدول يمكننا اشتقاق جدول التوزيع التكراري للنسبة ق .

جدول (١٣)

التوزيع التكراري للنسبة ق

التكرار	نسبة العينة ق
١	صفر
٦	٠,٥
٣	١
١٠	المجموع

ومن هذا الجدول يمكننا الحصول على التكرار النسبي بقسمة التكرار على مجموع التكرارات ، وبذلك نحصل على التوزيع الاحتمالي للنسبة ق . وهو يمثل توزيع معاينة للنسبة ق . وهو المبين في جدول (١٤)

جدول (١٤)

توزيع معاينة للنسبة ق

نسبة العينة (ق)	التكرار النسبي ح (ق)	ق ح (ق)	ق ^٢ ح (ق)
صفر	$0,1 = \frac{1}{10}$	صفر	صفر
٠,٥	$0,6 = \frac{6}{10}$	٠,٣	٠,١٥
١	$0,3 = \frac{3}{10}$	٠,٣	٠,٣٠
المجموع	١	٠,٦	٠,٤٥

ومن هذا الجدول يمكننا حساب الوسط الحسابي وتباين توزيع معاينة ق كما يلي :

$$\mu_1 = \text{مخ ق ح (ق)} = 0.6$$

$$\sigma_1^2 = \text{مخ ق}^2 \text{ ح (ق)} = 0.16$$

$$= 0.09 - (0.6)^2 = 0.09$$

$$\text{ويتضح مما سبق أن } \mu_1 = 0.6$$

وعلى وجه العموم ، يمكننا القول أن :

الوسط الحسابي لتوزيع معاينة النسبة ق يساوي نسبة المجتمع θ ، أي أن :

$$\mu_1 = \theta \quad (21)$$

كما أن تباين توزيع معاينة ق يمكن الحصول عليه من القانون :

$$\sigma_1^2 = \frac{(\theta - 1) \theta}{n} \quad (22)$$

ويكون الخطأ المعياري لتوزيع معاينة ق :

$$\sigma_1 = \sqrt{\frac{(\theta - 1) \theta}{n}} \quad (23)$$

ويستخدم قانوني (٢٢) ، (٢٣) عندما تكون نسبة حجم العينة ن إلى

$$\text{حجم المجتمع م هي : } \frac{n}{m} > 0.05$$

لما إذا كانت نسبة حجم العينة ن إلى حجم المجتمع م :

$$\frac{n}{m} \leq 0.05 \text{ ، فيجب استخدام معامل التصحيح ، أي أن :}$$

$$\text{عندما } \frac{n}{m} \leq 0.05 \text{ فإن :}$$

$$(٢٤) \quad \frac{\sqrt{\frac{n-p}{1-p}}}{\sqrt{\frac{(\theta-1)\theta}{n}}} = \sigma_q$$

وفي أغلب الأحوال فإن حجم العينة يكون صغيرا بالنسبة لحجم المجتمع ومن ثم فإن قانون (٢٣) هو المستخدم .

وكما سبق وبيننا بالنسبة لتوزيع معاينة \bar{p} ، فإن نظرية النهاية المركزية تنطبق على توزيع معاينة q .

نظرية ٤ :

تطبيقا لنظرية النهاية المركزية ، فعندما يكون حجم العينة كبيرا ، فإن توزيع معاينة النسبة q يكون معتدلا تقريبا ع $(\theta, \frac{(\theta-1)\theta}{n})$.
ويعتبر حجم العينة كبيرا إذا تحقق الشرطان :

$$n \leq \theta$$

$$n \leq (\theta-1)$$

مثال (٧) :

في أحد المجتمعات كانت نسبة المدخنين ٠,٣٥ ، فإذا سحبت عينة عشوائية من ١٠٠ مفردة فالمطلوب :

أولا : ما هو احتمال أن تكون نسبة المدخنين في هذه العينة أكبر من ٠,٤ ؟

ثانيا : ما هو احتمال أن تكون نسبة المدخنين في العينة في حدود ٠,٠٥ من نسبة المجتمع ؟

الحل :

$$\text{لولا : } \theta = 0,35 , (\theta - 1) = 0,65 , q = 0,40$$

يمكن تقريب توزيع معاينة النسبة ق إلى التوزيع المعتدل إذا تحقق

$$\text{الشرطان : } n \leq \theta \quad n \leq (\theta - 1)$$

$$n \leq 35 = 0,35 \times 100 = \theta$$

$$n \leq 65 = 0,65 \times 100 = (\theta - 1)$$

∴ يمكن تقريب توزيع معاينة النسبة ق إلى التوزيع المعتدل .

$$\text{حيث : } \mu = \theta = 0,35$$

$$\frac{0,65 \times 0,35}{100} \sqrt{\quad} = \frac{(\theta - 1) \theta}{n} \sqrt{\quad} = \sigma_q$$

$$0,048 =$$

$$Z = \frac{q - \theta}{\frac{(\theta - 1) \theta}{n} \sqrt{\quad}} = \frac{0,40 - 0,35}{0,048}$$

$$Z = \frac{0,05}{0,048}$$

$$Z = 1,04$$

$$1 - \Phi(1,04)$$

$$= 0,8508$$

$$= 0,1492$$

ثانياً : المطلوب أن تكون نسبة المخنن في العينة في حدود ٠,٠٥ من

نسبة المجتمع ، أي : $\theta \pm ٠,٠٥$

$٠,٠٥ \pm ٠,٣٥$

أي أن نسبة العينة تكون بين : ٠,٣٠ ، ٠,٤٠

$$\begin{aligned}
 & \text{ح } (٠,٣٠ \leq \theta \leq ٠,٤٠) \\
 & \left(\frac{٠,٣٥ - ٠,٤٠}{٠,٠٤٨} \geq \frac{\theta - \theta_0}{\frac{(\theta - 1)\theta}{n}} \geq \frac{٠,٣٥ - ٠,٣٠}{٠,٠٤٨} \right) \text{ح} = \\
 & \left(\frac{٠,٠٥}{٠,٠٤٨} \geq Z \geq \frac{٠,٠٥ - ٠,٣٠}{٠,٠٤٨} \right) \text{ح} = \\
 & \text{ح } (١,٠٤ \geq Z \geq ١,٠٤ -) = \\
 & (١,٠٤ -) \Phi - (١,٠٤) \Phi = \\
 & [(١,٠٤) \Phi - ١] - (١,٠٤) \Phi = \\
 & ١ - (١,٠٤) \Phi = \\
 & ١ - (٠,٨٥٠٨) = \\
 & ٠,١٤٩٢ =
 \end{aligned}$$

تمارين (١)

١ - في إحدى الشركات الصغيرة يقوم بالعمل ٤ موظفين فقط . وفيما

يلي الأجور الشهرية (بالجنديئات) لهؤلاء الموظفين :

٤٠٠ ، ٥٠٠ ، ٥٠٠ ، ٦٠٠

والمطلوب :

أولاً : حساب الوسط الحسابي والانحراف المعياري لهذا المجتمع

ثانياً : (أ) سحب جميع العينات الممكنة ذات الحجم ٢ في حالة

السحب بإرجاع ، ثم تكوين توزيع معاينة \bar{x} .

(ب) حساب الوسط الحسابي والانحراف المعياري لهذا

التوزيع .

ثالثاً : (أ) سحب جميع العينات الممكنة ذات الحجم ٢ في حالة

السحب بدون إرجاع ، ثم تكوين توزيع معاينة \bar{x} .

(ب) حساب الوسط الحسابي والانحراف المعياري لهذا

التوزيع .

٢ - فيما يلي مجتمع مكون من ٣ مفردات :

١٠ ، ١٥ ، ٢٧ ، ٣٠

والمطلوب :

أولاً : حساب الوسط الحسابي والانحراف المعياري لهذا المجتمع .

ثانياً : (أ) سحب جميع العينات الممكنة ذات الحجم ٣ في حالة

السحب بإرجاع ، ثم تكوين توزيع معاينة \bar{x} .

(ب) حساب الوسط الحسابي والانحراف المعياري لهذا

التوزيع .

ثالثاً : (أ) سحب جميع العينات الممكنة ذات الحجم ٣ في حالة السحب بدون إرجاع ، ثم تكوين توزيع معاينة \bar{S} .

(ب) حساب الوسط الحسابي والانحراف المعياري لهذا التوزيع .

٣ - كان عدد الموظفين في إحدى الشركات ١٠٠٠ موظف ، وكان الوسط الحسابي μ لأجور هؤلاء الموظفين هو ٨٠٠ جنيها شهريا بانحراف معياري (σ) ١٠٠ جنية . فإذا سحبت عينة عشوائية من هذا المجتمع وتم حساب الوسط الحسابي (\bar{S}) لأجر الموظف منها ، فالمطلوب حساب الوسط الحسابي لتوزيع معاينة \bar{S} ($\mu_{\bar{S}}$) والانحراف المعياري لهذا التوزيع ($\sigma_{\bar{S}}$) ، إذا كان حجم هذه العينة :

(أ) ٣٥ مفردة .

(ب) ٨٠ مفردة .

(ج) ٣٠٠ مفردة .

٤ - في أحد المجتمعات كان الوسط الحسابي $\mu = ١٢٠$ ، والانحراف المعياري $\sigma = ١٥$.

أولاً : إذا تم سحب عينة عشوائية من هذا المجتمع ، وكان الوسط الحسابي لتوزيع معاينة \bar{S} هو : $\mu_{\bar{S}} = ١٢٠$ ، والانحراف المعياري لهذا التوزيع هو : $\sigma_{\bar{S}} = ٣$. فإذا كانت العلاقة بين حجم العينة n وحجم المجتمع M هي : $\frac{n}{M} > ٠,٠٠٥$ ، فما هو حجم العينة ؟

ثانياً : إذا تم سحب عينة عشوائية من هذا المجتمع وكان الوسط الحسابي لتوزيع معاينة \bar{m} هو : $\mu = 120$ ، والانحراف المعياري لهذا التوزيع هو : $\sigma = 2$. فإذا كانت العلاقة بين حجم العينة n وحجم المجتمع m هي : $\frac{n}{m} > 0.05$ ، فما هو حجم العينة ؟

٥ - كان توزيع سرعة السيارات المسافرة على إحدى الطرق السريعة معتدلاً بوسط حسابي ٩٠ كم / الساعة وانحرافه المعياري ١٠ كم / الساعة . ولقد تم سحب عينة عشوائية من ٢٥ سيارة مسافرة على هذا الطريق ، فإذا كان \bar{m} هو الوسط الحسابي لسرعة السيارات في هذه العينة ، فالمطلوب :

أولاً : حساب الوسط الحسابي والانحراف المعياري لتوزيع معاينة \bar{m} .

ثانياً : ما هو شكل توزيع معاينة \bar{m} ؟ ولماذا ؟

٦ - كانت الطرود الواردة لأحد مكاتب البريد لها توزيع ملتسوي ناحية اليمين ، وكان وسطها الحسابي ٢,٥ كجم بانحراف معياري ٠,٩ كجم . وسحبت عينة عشوائية من ٤٠ طرد وارد لهذا المكتب . فإذا كان الوسط الحسابي لهذه العينة هو \bar{m} . فالمطلوب :

أولاً : حساب الوسط الحسابي والانحراف المعياري لتوزيع معاينة \bar{m} .

ثانياً : ما هو شكل توزيع معاينة \bar{m} ؟ ولماذا ؟

٧ - في أحد البنوك كانت أرصدة الحسابات الجارية ملتوية ناحية اليمين ، وكان وسطها الحسابي ٥٠٠ جنيهها وانحراف معياري ٢٠٠ جنيهها . فإذا سحبت عينة عشوائية من ٥٠ حساب جاري من هذا البنك ، وإذا كانت س هي الوسط الحسابي للأرصدة في هذه العينة . فالمطلوب :

أولا : حساب الوسط الحسابي والانحراف المعياري لتوزيع معاينة س .

ثانيا : ما هو شكل توزيع معاينة س ؟ ولماذا ؟

٨ - إذا كان حجم عبوات الأرز المعبأة في أحد المصانع يتوزع توزيعا معتدلا وسطه الحسابي ٥ كجم وانحرافه المعياري ٠,٠٥ كجم . وإذا سحبت عينة عشوائية من ٢٥ عبوة من إنتاج هذا المصنع ، فالمطلوب :

أولا : احتمال أن يكون الوسط الحسابي للعينة أقل من ٤,٨ كجم

ثانيا : احتمال أن يكون الوسط الحسابي للعينة أكبر من ٤,٩ كجم

ثالثا : احتمال أن ينحصر الوسط الحسابي للعينة بين ٤,٦ كجم و ٤,٧ كجم .

٩ - في أحد المجتمعات كان الوسط الحسابي لدخل الفرد السنوي هو ٦٠٠٠ جنيهها ، بانحراف معياري ٢٠٠٠ جنيهها ، وكان توزيع الدخل ملتويا جهة اليمين . فإذا سحبت عينة عشوائية من ٤٠٠ فرد ، فالمطلوب إيجاد احتمال أن يكون الوسط الحسابي لهذه العينة

أولا : أقل من ٤٥٠٠ جنيهها .

ثانيا : بين ٤٥٠٠ جنيهها و ٧٥٠٠ جنيهها .

ثالثاً : في حدود ١٠٠٠ جنيه من متوسط المجتمع .

رابعاً : أقل من متوسط المجتمع بمبلغ ٥٠٠ جنيه أو أكثر .

١٠ - في أحد المدن كان توزيع فواتير الكهرباء توزيعاً ملتوياً ، وكان

وسطه الحسابي ٧٠ جنيهها بانحراف معياري ٣٠ جنيه . فإذا سحبت

عينة عشوائية من ١٠٠ أسرة من هذه المدينة ، فالمطلوب إيجاد

احتمال أن يكون الوسط الحسابي لهذه العينة

أولاً : أكثر من ٧٥ جنيهها .

ثانياً : بين ٥٣ جنيهها و ٥٨ جنيهها .

ثالثاً : في حدود ٨ جنيهات من الوسط الحسابي للمجتمع .

رابعاً : أكثر من متوسط المجتمع بمبلغ ١٠ جنيهات على الأقل .

١١ - ينتج أحد المصانع مصابيح كهربائية ، ومن المعروف أن

الانحراف المعياري لعمر هذه المصابيح هو ١٢٠ ساعة . ولقد

سحبت عينة عشوائية من ١٠٠ مفردة ووجد أن الوسط الحسابي

لعمر المصابيح في هذه العينة هو ١٥٠٠ ساعة . والمطلوب : ما

هو احتمال أن يكون عمر المصباح في هذه العينة في حدود

٢٠ ساعة من متوسط عمر المصابيح في هذا المصنع ؟

١٢ - في إحدى المحافظات كانت نسبة المدرسات (الإناث) في

المدارس الحكومية ٦٠ ٪ . فإذا سحبت عينة عشوائية من ٨٠ من

مدرسين هذه المدارس ، وكانت نسبة المدرسات (الإناث) في هذه

العينة ق ، فالمطلوب :

أولاً : حساب الوسط الحسابي والانحراف المعياري لتوزيع ق

ثانياً : ما هو شكل توزيع معاينة ق ؟

١٣ - في إحدى الامتحانات كانت نسبة الناجحين ٨٥ ٪ . وسحبت عينة عشوائية من ٥٠ طالب . فإذا كانت ق هي نسبة الناجحين في هذه العينة فالمطلوب :

أولا : حساب الوسط الحسابي والانحراف المعياري لتوزيع ق
ثانيا : ما هو شكل توزيع معاينة ق ؟

١٤ - سجلت إدارة المرور في إحدى المحافظات أن نسبة السيدات اللاتي يملكن رخص قيادة هي ٤٠ ٪ . ولقد سحبت عينة من ١٠٠ رخصة قيادة وكانت نسبة السيدات فيها هي ق . والمطلوب إيجاد احتمال أن تكون ق :

أولا : أقل من ٠,٣٨

ثانيا : تتحصر بين ٠,٣٧ و ٠,٤٣

ثالثا : في حدود ٠,٠٥ من نسبة المجتمع .

١٥ - يقوم أحد المصانع بإنتاج بطاريات للسيارات . وتدعي الإدارة أن ٩٠ ٪ من إنتاج المصنع مطابق للمواصفات . فإذا سحبت عينة عشوائية من ١٠٠ بطارية من إنتاج هذا المصنع وكانت نسبة البطاريات المطابقة للمواصفات هي ق ، فالمطلوب :

أولا : ما هو احتمال أن تكون نسبة العينة في حدود ٠,٠٥ من نسبة المجتمع ؟

ثانيا : ما هو احتمال أن تكون نسبة العينة أقل من نسبة المجتمع بمقدار ٠,٠٢ أو أكثر ؟

ثالثا : ما هو احتمال أن تكون نسبة العينة أكبر من نسبة المجتمع بمقدار ٠,٠٣ أو أكثر ؟

١٦ - من المعروف أن نسبة التالف من إنتاج أحد الآلات هو ٧ ٪ .
ويقوم مفتش الرقابة على الجودة كل أسبوع بسحب عينة عشوائية
من ١٠٠ منتج من إنتاج هذه الآلة . فإذا كانت نسبة التالف ٩ ٪
أو أكثر ، فإن المفتش يقرر إيقاف الإنتاج وتصليح الآلة .
والمطلوب : ما هو احتمال أن يقرر المفتش إيقاف الإنتاج وتصليح
الآلة بعد سحب عينة من ١٠٠ مفردة ؟

الفصل الثاني

تقدير معالم المجتمع

Estimation of the Population Parameters

مقدمة :

سبق وعرفنا في الفصل السابق المجتمع بأنه " جميع المفردات محل الدراسة سواء كانت في شكل إنسان أو حيوان أو جناد أو أشياء غير ملموسة ". كما عرفنا العينة بأنها " مجموعة من مفردات المجتمع " . وبينما أن أي مقياس إحصائي في المجتمع يسمى " معلمة " وأن أي مقياس إحصائي في العينة يسمى " إحصائية " .

ولقد سبق وبيننا أن دراسة جميع مفردات المجتمع — والمسماة بالحصص الشامل — تتطلب تكاليف باهظة وتستهلك كثيراً من الوقت والجهد ، وقد يؤدي هذا الأسلوب إلى تلف الوحدات محل الدراسة . وبالإضافة إلى هذه الصعوبات ، فإن الحصول على معالم المجتمع من الأمور التي يتعذر الوصول إليها إن لم يكن ذلك مستحيلاً . لذلك لجأ الإحصائيون إلى استخدام المقاييس الإحصائية الناتجة من العينة ، أي " إحصائيات " العينة للتعرف على معالم المجتمع ، وهذا ما يسمى بالاستدلال الإحصائي Statistical Inference . وعندما نذكر العينة هنا نعني العينة العشوائية . ومن ثم يمكن تعريف الاستدلال الإحصائي بوصفه " مجموعة الأساليب الإحصائية التي بمقتضاها يمكننا الاستدلال على معالم المجتمع باستخدام إحصائيات عينة عشوائية من هذا المجتمع " .

وينقسم الاستدلال الإحصائي إلى موضوعين : تقدير معالم المجتمع واختبارات الفروض الإحصائية . ويتناول هذا الفصل دراسة تقدير معالم المجتمع بينما يختص الفصل الثالث بدراسة اختبارات الفروض الإحصائية . وهناك أسلوبان لتقدير معالم المجتمع ، التقدير بنقطة Point estimation والتقدير بفترة Interval estimation . وفي دراستنا لتقدير معالم المجتمع في الفصل الحالي سنتعرض للمباحث الآتية : التقدير بنقطة ، والتقدير بفترة ، ثم تقدير فترة ثقة لمتوسط مجتمع في حالة العينات الكبيرة ، وتقدير فترة ثقة لمتوسط مجتمع في حالة العينات الصغيرة ، فتقدير فترة ثقة لنسبة مجتمع ، فتقدير حجم العينة لتقدير متوسط مجتمع ، فتقدير حجم العينة لتقدير نسبة مجتمع .

(٢ - ١) أسلوب التقدير بنقطة :

وكما وضعنا في تعريف الاستدلال الإحصائي فإن عملية تقدير معالم المجتمع تتم عن طريق إحصائيات العينة ، فمثلاً إذا أردنا معرفة متوسط أعمار الطلبة في كلية التجارة يمكن أخذ عينة من ١٠٠ طالب وحساب الوسط الحسابي لعمرهم والانحراف المعياري لهذا العمر ، وليكن مثلاً الوسط الحسابي لأعمارهم $\bar{x} = 20$ سنة وانحراف معياري $s = 3$ سنوات ، فنقول هنا أن القيمة ٢٠ سنة هي تقدير نقطة للوسط الحسابي للمجتمع μ وأن القيمة ٣ سنوات هي تقدير نقطة للانحراف المعياري للمجتمع σ . ومن ثم فإن قيمة إحصائية العينة هي تقدير نقطة لمعلمة المجتمع . ومن ناحية أخرى فإن الإحصائية المستخدمة لتقدير معلمة المجتمع — والتي يمكن للتعبير عنها بصيغة رياضية تبين الطريقة التي يتم بها حساب تقدير النقطة — تسمى "مقدر النقطة" Point estimator . فمثلاً للوسط الحسابي \bar{x} هو مقدر نقطة للوسط الحسابي للمجتمع μ ، حيث : $\bar{x} = \frac{\sum x}{n}$.

وبالمثل فإن الانحراف المعياري في العينة σ هو مقدار نقطة للانحراف المعياري للمجتمع σ ، والنسبة في العينة q هي مقدار نقطة لنسبة المجتمع θ .
 ولقد جرى العرف على استخدام علامة $(^A \text{ هات})$ فوق رمز المعلمة للدلالة على مقدار النقطة فمثلاً إذا أردنا الدلالة على مقدار نقطة لمعلمة المجتمع μ ، فإننا نكتب $\hat{\mu}$. بمعنى آخر فإن $(^A \text{ هات})$ ترمز لمقدار المعلمة التي تكتب تحتها . وبالمثل فإن :

$\hat{\mu}$ هي مقدار نقطة للمعلمة μ .

$\hat{\sigma}$ هي مقدار نقطة للمعلمة σ .

$\hat{\theta}$ هي مقدار نقطة للمعلمة θ . وهكذا

وللتخيص ما جاء في هذا المبحث نأخذ المثال السابق الخاص بأعمار طلاب كلية التجارة ، ومنه نجد أن :

اسم المعلمة	رمز المعلمة	مقدار نقطة	تقدير نقطة
الوسط الحسابي	μ	$\hat{\mu} = \overline{X}$	$\hat{\mu} = 20$ سنة
الانحراف المعياري	σ	$\hat{\sigma} = s$	$\hat{\sigma} = 3$ سنوات

ومن الجدير بالذكر أن مقدار النقطة هو متغير عشوائي له توزيع احتمالي هو توزيع للمعينة الخاصة به ، بينما تقدير النقطة هو مقدار ثابت .

وعند اختيار مقدار نقطة ما ، يثار التساؤل عما إذا كان هذا المقدار " جيد " أم لا . فعلى سبيل المثال : هل الوسط الحسابي مقدار جيد للمعلمة μ ؟ أم هل من الأفضل استخدام الوسيط ؟ للإجابة على هذه التساؤلات يجب الأخذ في الاعتبار بعض المعايير وهي : عدم التحيز ، والكفاءة ، والكفاية ، والاتساق ، ومتوسط مربع الخطأ .

١ - عدم التحيز : Unbiasedness

يقال أن مقدر نقطة هو مقدر غير متحيز لمعلمة المجتمع إذا كان متوسط الإحصائية - المحسوبة من جميع العينات العشوائية الممكنة المحسوبة من هذا المجتمع والتي لها نفس الحجم - يساوي معلمة المجتمع . أو بعبارة أخرى فإن مقدر نقطة يقال عنه أنه غير متحيز إذا كان متوسط توزيعه العيني يساوي معلمة المجتمع . ومن ثم فإن $\hat{\mu}$ هو مقدر نقطة غير متحيز للمعلمة μ إذا كان :

(١)

$$E(\hat{\mu}) = \mu$$

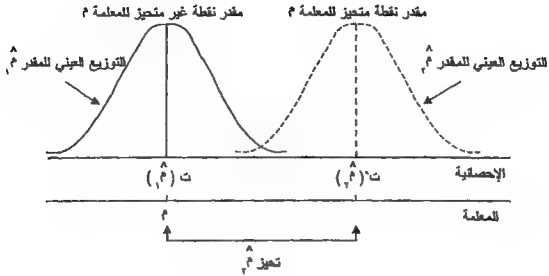
وإذا كانت $\hat{\mu}$ مقدر نقطة متحيز ، فإن مقدار التحيز (أ) يقاس كما يلي :

(٢)

$$A = E(\hat{\mu}) - \mu$$

وبين شكل (١) مقدر نقطة : $\hat{\mu}_1$ مقدر نقطة غير متحيز ، $\hat{\mu}_2$ مقدر نقطة متحيز . ويوضح الرسم أن $\hat{\mu}_1$ تعطي تقديرات بعيدة عن المعلمة μ ، فهي حين أن $\hat{\mu}_1$ تعطي تقديرات قريبة من المعلمة μ . ويعتبر المقدر $\hat{\mu}_2$ متحيز لأن $E(\hat{\mu}_2) \neq \mu$. ويقاس التحيز بالمقدار أ وهو الفرق بين $E(\hat{\mu}_2)$ و μ أي أن :

$$A = E(\hat{\mu}_2) - \mu$$



شكل (١) : توزيعي معاينة مقدر غير متحيز وآخر متحيز

وقد يكون المقدر المتحيز مقدراً مرغوباً فيه إذا كان مقدار التحيز صغيراً ، طالما أن هذا المقدر يتمتع بخصائص أخرى مرغوب فيها .

ولقد تبين من دراستنا في المعاينة الإحصائية أن الوسط الحسابي لتوزيع المعاينة للمتغير العشوائي يساوي الوسط الحسابي للمجتمع μ . ويمكن استخدام الرمز \bar{S} - أي توقع \bar{S} - للدلالة على الوسط الحسابي لتوزيع المعاينة للمتغير العشوائي \bar{S} . أي أن :

$$E(\bar{S}) = \mu$$

إذن الوسط الحسابي للعينة (\bar{S}) يعتبر مقدر غير متحيز للوسط الحسابي للمجتمع μ .

أما بالنسبة للوسيط ، فإذا كان للتوزيع ملتوياً فإن الوسيط المحسوب من العينة يعتبر تقديراً متحيزاً للوسط الحسابي μ ، أي أن : $E(\text{الوسيط}) \neq \mu$. فعلى سبيل المثال الوسيط المحسوب من عينة عشوائية لدخل مجموعة من

الأسر يكون أقل بكثير من الوسط الحسابي لدخل الأسرة في المجتمع ، وهذا لأنه عادة يكون توزيع الدخل ملتوياً ناحية اليمين .

أما بالنسبة للتباين للمأخوذ من العينة والذي يحسب من المعادلة :

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

فهي تعتبر مقدار نقطة غير متحيز للمعلمة σ^2 لأن :

$$E(s^2) = \sigma^2$$

وبلاحظ أننا قسمنا هنا على $n-1$ (وهي درجات الحرية)^(١) بدلاً

من القسمة على n .

أما الانحراف المعياري للعينة فهو مقدار نقطة متحيز للمعلمة σ ، وذلك لأن :

$$E(s) \neq \sigma$$

٢ - الكفاءة : Efficiency

نقاس كفاءة مقدار غير متحيز عن طريق تباين توزيعه العيني . فإذا

كان لدينا مقدري نقطة غير متحيزين من عينتين لهما نفس الحجم ، فإن مقدر

النقطة الأعلى كفاءة نسبياً هو ذلك المقدر ذو التباين الأصغر .

^(١) يمكن تعريف درجات الحرية بأنها عدد المشاهدات التي يمكن اختيارها بحرية ، أو عدد المتغيرات التي يمكن أن تتغير بحرية ، أو عدد المتغيرات المستقلة . ففي حالة وجود مجموع مربعات كميات معينة ، فإن درجات الحرية تعرف بأنها عدد المربعات ناقص عدد المتغيرات المستقلة المفروضة على الكميات محل البحث فعدد n من المشاهدات لدينا n من مربعات الانحرافات التوزع عن وسطها الحسابي ، ولكن هناك $n-1$ فقط منها هي المستقلة ، بمعنى أنه إذا عرفنا $n-1$ من هذه الانحرافات نستطيع تحديد الانحراف التوزعي . والمعب يرجع إلى أن لدينا قيداً وهو : أن مجموع الانحرافات عن الوسط الحسابي يجب أن يساوي صفر . ومن ثم فإن مجموع مربعات الانحرافات n من القوس عن وسطها الحسابي لها $n-1$ درجات حرية .

ومن ثم ، فإذا كان لدينا مقدرتي نقطة غير متحيزين \hat{m}_1 ، \hat{m}_2 للمعلمة m ، فإن \hat{m}_1 يكون أكثر كفاءة نسبياً إذا كان :

$$(3) \quad \sigma^2(\hat{m}_1) < \sigma^2(\hat{m}_2) \quad \text{حيث} \quad t(\hat{m}_1) = t(\hat{m}_2) = m$$

مثال (١) :

المطلوب مقارنة كلا من الوسيط الحسابي والوسيط من حيث الكفاءة كمقدرتي نقطة غير متحيزين للوسيط الحسابي لمجتمع يتوزع توزيعاً معتمداً ، علماً بأن كلا من الوسيط الحسابي والوسيط غير متحيزين وأن :

$$\begin{aligned} \text{تباين } \bar{m} &= \frac{\sigma^2}{n} \\ \text{تباين الوسيط} &= \frac{\sigma^2}{n} \cdot 1,57 \end{aligned}$$

تباين الوسيط أكبر من تباين الوسيط الحسابي ، مع افتراض نفس حجم العينة . ومن ثم فإن الوسيط الحسابي أكثر كفاءة من الوسيط . ويمكن قياس الكفاءة النسبية لمقدر ما بالنسبة لمقدر آخر عن طريق النسبة بين تبايني المقدرين . فإذا كان المقدرين غير متحيزين فإن :

(٤)

$$\frac{\text{تباين } \hat{m}_1}{\text{تباين } \hat{m}_2} = \text{الكفاءة النسبية}$$

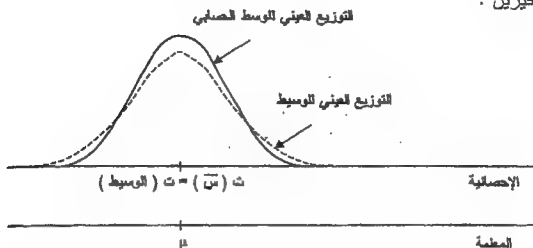
وفي مثالنا السابق :

$$\begin{aligned} \frac{\text{تباين } \bar{m}}{\text{تباين الوسيط}} &= \text{الكفاءة النسبية} \\ &= \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n} \cdot 1,57} = 0,64 \end{aligned}$$

وبما أن هذه النسبة أقل من الواحد الصحيح ، إذن تباين \bar{s} أقل من تباين الوسيط ومن ثم فإن لوسط الحسابي أكثر كفاءة من الوسيط .

ويوضح شكل (٢) هذه الحالة . ومن الملاحظ أن المقدريين غير

متحيزين .



شكل (٢) : التوزيع العيني للوسط الحسابي مقارن بالتوزيع العيني للوسيط من حيث الكفاءة بوصفهما مقدري نقطة غير متحيزين للمعلمة μ

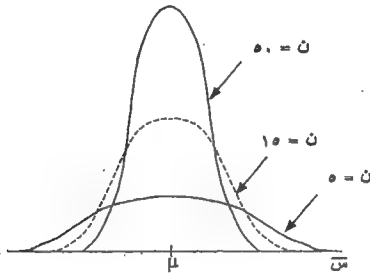
ولكن التوزيع العيني للوسط الحسابي أكثر تركزا حول μ عن التوزيع العيني للوسيط

٣ - الاتساق : Consistency

يقال عن إحصائية أنها مقدر متسق إذا اقترب المقدر من معلمة المجتمع مع ازدياد حجم العينة . ولقد رأينا أن $\frac{\sigma^2}{n} = \sigma_{\bar{y}}^2$. وهذا يعني أنه بازياد حجم العينة ن فإن تباين توزيع معاينة \bar{s} ($\sigma_{\bar{y}}^2$) سينخفض ، أي أن الأوساط الحسابية لتمينات (\bar{s}_1 ، \bar{s}_2 ، ... ، \bar{s}_n) تقترب من الوسيط الحسابي للمجتمع μ مع ازدياد حجم للعينة . ومن ثم فإن الوسيط الحسابي \bar{s} هو مقدر متسق للمعلمة μ .

ويكون المقدّر متسقاً إذا كان نهايته يقترب من الصفر عندما يقترب حجم العينة من ∞ . وهنا أيضاً نجد أن الوسط الحسابي \bar{x} يحقق هذا الشرط لأن $\frac{\sigma^2}{n}$ تقترب من الصفر عندما n تقترب من ∞ . ويمكن أيضاً تبيان أن \bar{x} هو مقدّر متسق للمعلمة σ^2 .

وبين شكل (٣) أن \bar{x} هو مقدّر متسق للمعلمة μ عندما يكون التوزيع العيني معتدل ، وبين الشكل أن \bar{x} تزداد اقتراباً من μ كلما ازداد حجم العينة .



شكل (٣) : \bar{x} مقدّر متسق للمعلمة μ عندما يكون التوزيع العيني معتدل

٤ - الكفاية : Sufficiency

يقال عن مقدّر أنه كافٍ إذا استخدم جميع البيانات الموجودة في العينة والخاصة بحساب المعلمة المراد تقديرها . وإذا كان هناك مقدّر كافٍ فلا يكون مجدياً استخدام أي مقدّر آخر أقل كفاية ، فالمقدّر الكاف يستخدم كل المعلومات الموجودة في العينة والخاصة بتقدير المعلمة . ويعتبر الوسط الحسابي \bar{x} مقدّر

كاف للمعلمة μ لأنه يستخدم في حسابيه نفس المعلومات التي تستخدمها المعلمة μ . فلحساب μ نجمع جميع القيم ونقسم على عددها . ونفعل نفس الشيء في العينة لحساب \bar{x} . أما الوسيط فهو لا يعتبر مقدار كاف للمعلمة μ ، فلحساب الوسيط نوجد ترتيب الوسيط ، ثم نوجد القيمة الوسطى ، ولا نستخدم جميع القيم كما يحدث عند حساب μ . هذا وتعتبر النسبة q مقدار كاف للمعلمة θ لأنها تستخدم في حسابها نفس المعلومات التي تستخدمها المعلمة θ . فلحساب النسبة θ في المجتمع نقسم عدد المفردات التي لها نفس الصفة على عدد المفردات الكلية . ونفعل نفس الشيء في العينة لحساب النسبة q .

٥ - متوسط مربع الخطأ : Mean squared error

يمزج هذا للمعيار معيار عدم التحيز ومعيار الكفاءة ، وهذا المعيار مفيد في حالة مقارنة مقدرين أحدهما أو كلاهما متحيز . ويمزج متوسط مربع الخطأ للمقدر $\hat{\mu}$ بتباين التوزيع العيني للمقدر $\hat{\mu}$ أي $\sigma^2(\hat{\mu})$ وتحيز المقدر أي $t(\hat{\mu}) - \hat{\mu}$.

ويمكن تعريف متوسط مربع الخطأ للمقدر μ كالآتي :

$$(٥) \quad \text{متوسط مربع الخطأ} = \sigma^2(\hat{\mu}) + [t(\hat{\mu}) - \hat{\mu}]^2$$

وعند مقارنة مقدرين ، فإن المقدر الذي له أقل متوسط مربع الخطأ يقال عنه أنه ذو كفاءة نسبية أعلى في متوسط مربع الخطأ عن المقدر الآخر .

ووفقا لهذا المعيار يفضل المقدر المتحيز الذي يتمتع بتوزيع عيني مركز حول المعلمة μ ، على مقدر غير متحيز له توزيع عيني ذو تشتت أكبر .

مثال (٢) : فإذا كان لدينا المقدرين الآتين :

المقدر	التباين	التحيز	متوسط مربع الخطأ
$\hat{\mu}_1$	$\sigma^2 (\hat{\mu})$	$\bar{t} - (\hat{\mu})$	$\sigma^2 (\hat{\mu}) + [t - (\hat{\mu})]^2$
$\hat{\mu}_1$	١٠	٣	$١٠ + ٣^2 = ١٩$
$\hat{\mu}_2$	٤٠	٠	$٤٠ + ٠ = ٤٠$

وطبقا لمعيار متوسط مربع الخطأ يفضل المقدر $\hat{\mu}_1$ على $\hat{\mu}_2$.

وبالإضافة إلى المعايير السابقة فإنه هناك معيارين آخرين لمعرفة إذا كان المقدر جيدا أم لا ، وهما : طريقة الإمكان الأكبر Maximum Likelihood Method وطريقة المربعات الصغرى Method of Least Squares . وإن نتعرض لهما في دراستنا .

(٢ - ٢) أسلوب التقدير بفترة :

من دراستنا للتقدير بنقطة تبين لنا أن الوسط الحسابي \bar{x} المحسوب من العينة هو مقدر جيد لمعلمة المجتمع μ . ولكن هذا لا يعني أن \bar{x} تماثلي μ بالضبط ، فقد تأخذ \bar{x} قيم أقل من أو أكبر من μ طبقا للعينة المحسوبة منها . ومن ثم فإن التقدير بنقطة يعطي قيمة لمعلمة المجتمع تكون في أغلب الأحيان مختلفة تماما عن القيمة الحقيقية لهذه المعلمة . لذلك يكون من الأفضل استخدام أسلوب التقدير بفترة . وطبقا لهذا الأسلوب فإنه يتم وضع فترة حول مقدر النقطة بحيث يكون من المحتمل أن تحتوي هذه الفترة على معلمة المجتمع باحتمال محدد مقدما وهو ما يسمى بدرجة الثقة degree of confidence أو مستوى الثقة confidence level . ويمكننا الحصول على هذه الفترة لتحقيق

أي درجة ثقة مطلوبة لذلك سميت هذه الفترة بفترة الثقة confidence interval كما سمي حدي هذه الفترة بحدي الثقة confidence limits . ولإيضاح ذلك سنأخذ الحالة القصوى ، فيمكننا القول بدرجة ثقة ١٠٠ ٪ بأن الفترة من $-\infty$ إلى $+\infty$ تشمل على معلمة المجتمع ، وهذه الفترة ليس لها في الواقع أهمية عملية . ويمكننا تضيق هذه الفترة ولكن هذا يتم ألا وهو تخفيض درجة الثقة بأن هذه الفترة تحتوي على المعلمة المجهولة .

وجملة القول : كل فترة مصحوبة بدرجة ثقة معينة لذلك سميت بفترة الثقة . وتحدد درجة الثقة — المصاحبة لفترة الثقة — مدى الثقة التي تكون لدينا — بأن هذه الفترة تحتوي على معلمة المجتمع الحقيقية . ويرمز لدرجة الثقة بالرمز ١٠٠ ($1 - \alpha$) ٪ ، حيث α هي الحرف الإغريقي ويقرأ " ألفا " ، وتسمى α بمستوى المعنوية significance level ، كما يسمى الاحتمال ($1 - \alpha$) بمعامل الثقة confidence coefficient وعادة يتم استخدام درجات الثقة ٩٠ ٪ أو ٩٥ ٪ أو ٩٩ ٪ ، فعلى سبيل المثال إذا قلنا أن :

$$ح (30 \leq \mu \leq 50) = 0.99$$

فإن هذا يعني أنه باحتمال قدره ٩٩ ٪ تحتوي الفترة ما بين ٣٠ ، ٥٠ على القيمة الحقيقية لمتوسط المجتمع μ . ولكن من الخطأ القول بأنه باحتمال ٩٩ ٪ تنحصر μ بين ٣٠ و ٥٠ وذلك لأن معلمة المجتمع قيمة ثابتة والاحتمالات تتعلق دائماً بالمتغيرات العشوائية ، فالاحتمال هنا هو معامل الثقة (٩٩ ٪) بينما للمتغيرات العشوائية هي حدي الثقة (٣٠ ، ٥٠) .

هذا ويفضل بعض الإحصائيون مناقشة تقدير معالم المجتمع والفروض الإحصائية بمعلومية أو عدم معلومية σ . ولكن في دراستنا هنا يتم استخدام معيار العينات الكبيرة والعينات الصغيرة ، والسبب في ذلك يرجع إلى أن الانحراف المعياري للمجتمع يكون في غالب الأحيان دائماً مجهولاً . لذلك

فدراسة تقدير معالم المجتمع واختبارات الفروض ، طبقاً لما إذا كانت العينات كبيرة أم صغيرة ، تكون أكثر واقعية من كون σ معلومة أو غير معلومة .

(٢ - ٣) تقدير فترة ثقة لمتوسط مجتمع في حالة العينات الكبيرة :

سبق وذكرنا عند دراستنا للمعينة الإحصائية أنه إذا سحبنا عينة عشوائية حجمها n من مجتمع معين وحسبنا الوسط الحسابي منها \bar{x} ، ثم سحبنا عينة عشوائية أخرى من نفس المجتمع حجمها n وحسبنا منها الوسط الحسابي \bar{y} وهكذا . . . إلى أن يتم سحب جميع العينات العشوائية الممكنة التي لها نفس الحجم n من هذا المجتمع ، نجد أن \bar{x} متغير عشوائي له توزيع احتمالي هو التوزيع العيني للوسط الحسابي \bar{x} . وطبقاً لنظرية النهاية المركزية ، فعندما تكون العينة كبيرة فإن التوزيع العيني للوسط الحسابي \bar{x} يتبع تقريباً التوزيع المعتدل الذي وسطه الحسابي μ وتباينه $\frac{\sigma^2}{n}$ ، وهذا بصرف النظر عن شكل التوزيع الأصلي للمجتمع . وتعتبر العينة كبيرة إذا كان حجمها $n \leq 30$.

ويجب التفرقة هنا بين حالة ما إذا كانت σ معلومة أم غير معلومة .

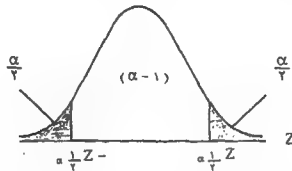
أ - إذا كانت σ معلومة :

وبما أن التوزيع العيني يتبع للتوزيع المعتدل $E(\mu, \frac{\sigma^2}{n})$ ، ومن ثم فإن المتغير العشوائي Z يتبع للتوزيع المعتدل المعياري بوسط حسابي صفر وانحراف معياري ١ ، أي $E(1, 0)$ ، حيث :

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

حيث $\frac{\sigma}{\sqrt{n}}$ هو الخطأ المعياري .

فإذا تم اختيار درجة ثقة ١٠٠ ($\alpha - 1$) % ، فهذا يعني أن معامل الثقة (أي الاحتمال) هو ($\alpha - 1$) . ومن ثم فإن المساحة الباقية تحت المنحنى [$\alpha = (\alpha - 1) - 1$] تقسم بالتساوي بين طرفي التوزيع أي ($\frac{\alpha}{2}$) . وكما هو موضح في شكل (٤) فإن ($\alpha - 1$) من المساحة الكلية تقع بين $-\alpha \frac{1}{2} Z$ ، $\alpha \frac{1}{2} Z$. حيث $\alpha \frac{1}{2} Z$ هي قيمة Z التي تكون المساحة إلى يمينها مساوية $\alpha \frac{1}{2}$. وتستخرج هذه القيمة من جدول التوزيع المعتدل المعياري ، وبما أن هذه الجداول تعطي فقط قيم Z للموجبة ، فإننا سترمز إلى قيمة Z التي تكون المساحة على يسارها مساوية $\alpha \frac{1}{2}$ بالرمز $-\alpha \frac{1}{2} Z$. وبالتالي فإن :



شكل (٤) : التوزيع المعتدل المعياري مبيناً حدي الثقة عليه

$$\alpha - 1 = P(-\alpha \frac{1}{2} Z \leq Z \leq \alpha \frac{1}{2} Z) \quad \text{ح}$$

$$(٦) \quad \alpha - 1 = P(-\alpha \frac{1}{2} Z \geq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \alpha \frac{1}{2} Z) \quad \text{ح}$$

بضرب طرفي المتباينة في $\frac{\sigma}{\sqrt{n}}$ (حيث $\frac{\sigma}{\sqrt{n}}$ مقدار موجب) ،

$$\alpha - 1 = P(-\alpha \frac{1}{2} Z \geq \bar{X} - \mu \geq \alpha \frac{1}{2} Z) \quad \text{ح}$$

ويطرح \bar{s} من طرفي المتباينة :

$$\alpha - 1 = \left(\frac{\sigma}{\sqrt{n}} \alpha \frac{1}{2} Z + \bar{s} - \mu \geq \mu \geq \frac{\sigma}{\sqrt{n}} \alpha \frac{1}{2} Z - \bar{s} \right) \text{ ح}$$

ويضرب طرفي المتباينة في -1 :

$$(V) \quad \alpha - 1 = \left(\frac{\sigma}{\sqrt{n}} \alpha \frac{1}{2} Z - \bar{s} \leq \mu \leq \frac{\sigma}{\sqrt{n}} \alpha \frac{1}{2} Z + \bar{s} \right) \text{ ح}$$

وهذا يعني أن هناك احتمالاً قدره $(\alpha - 1)$ بأن الفترة التي يكون

$$\frac{\sigma}{\sqrt{n}} \alpha \frac{1}{2} Z - \bar{s} = \text{حدها الأدنى} , \quad \frac{\sigma}{\sqrt{n}} \alpha \frac{1}{2} Z + \bar{s} = \text{حدها الأعلى}$$

تحتوي على قيمة المعلمة μ

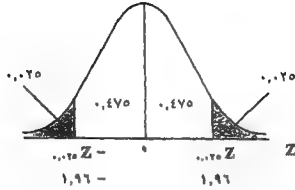
ويسمى المقدار $\left(\frac{\sigma}{\sqrt{n}} \alpha \frac{1}{2} Z \right)$ بخطأ التقدير Error of estimate ^(١).

وجملة القول :

$\frac{\sigma}{\sqrt{n}} \alpha \frac{1}{2} Z \pm \bar{s}$ <p>(٨)</p>	<p>هو مقدار فترة ثقة ١٠٠ ($\alpha - 1$) % لمتوسط مجتمع معلوم التباين</p> <p>عندما $n \leq 30$</p>
---	---

لإيضاح كيفية الحصول على قيمة $\alpha \frac{1}{2} Z$ من جدول التوزيع المعتدل المعياري ، فعلى سبيل المثال إذا كان المطلوب تحديد فترة ثقة ٩٥ % للمعلمة μ ، فهذا يعني أن المساحة تحت المنحنى المعتدل بين نقطتين هي ٩٥ % أي ٠,٠٥ على يمين ويسار μ ، ومن ثم ٠,٢٥ على كل من يمين ويسار μ ، كما هو مبين في شكل (٥) .

^(١) يجب ملاحظة أن كلمة خطأ هنا لا تعني المعنى المعروف لها . ولكن يقصد بها الاختلافات في قيم الإحصائيات من عينة إلى أخرى .



شكل (٥) : إيجاد قيمة Z ٠,٠٢٥

إذا أخذنا جدول المنحني المعتدل المعياري الذي يعطي المساحة من $-\infty$ إلى Z ، نتبع الخطوات التالية :

١- نوجد قيمة α ، بما أن $\alpha - 1 = 0.95$ ، إذن $\alpha = 0.05$

٢- نوجد قيمة $\frac{1}{\alpha}$ ، أي $\frac{1}{0.05} = 20$

٣- نوجد $(\frac{1}{\alpha} - 1)$ ، أي $20 - 1 = 19$

٤- نبحث في الجدول نفسه أي في المساحة حتى نجد قيمة ٠,٩٧٥ ، فنجد أن هذه القيمة تقع أمام $Z = 1.9$ وتحت $Z = 0.6$ وبالتالي قيمة Z تساوي ١,٩٦ .

ويبين الجدول التالي قيم Z و $\frac{1}{\alpha}$ لدرجات ثقة مختلفة شائعة الاستخدام :

درجة الثقة ١٠٠ ($\alpha - 1$) %	$\alpha - 1$	α	$\frac{1}{\alpha}$	$\alpha \div Z$
٩٠ %	٠,٩٠	٠,١٠	١٠	١,٦٤٥
٩٥ %	٠,٩٥	٠,٠٥	٢٠	١,٩٦٠
٩٩ %	٠,٩٩	٠,٠١	١٠٠	٢,٥٧٥

مثال (٣) :

سحبت عينة عشوائية من ١٠٠ مصباح كهربائي من إنتاج أحد المصانع فوجد أن الوسط الحسابي لعمر المصباح ١٠٠٠ ساعة . فإذا علمت أن الانحراف المعياري لعمر المصباح في المجتمع هو ١٥٠ ساعة ، فالمطلوب :

(أ) إيجاد تقدير نقطة لمتوسط عمر المصباح .

(ب) إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط عمر المصباح في هذا المصنع .

(ح) إيجاد تقدير فترة ثقة ٩٩ ٪ لمتوسط عمر المصباح في هذا المصنع .

(د) إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط عمر المصباح إذا كان حجم العينة ٢٠٠ مصباح .

الحل :

$$n = 100, \quad \bar{x} = 1000 \text{ ساعة}, \quad \sigma = 150 \text{ ساعة}, \quad 1 - \alpha = 0.95, \\ \alpha = 0.05, \quad \frac{1}{\sqrt{\alpha}} Z_{0.025} = 1.96$$

(أ) تقدير نقطة لمتوسط عمر المصباح : $\hat{\mu} = \bar{x} = 1000 \text{ ساعة} .$

(ب) بما أن $n < 30$ فيمكننا تطبيق نظرية النهاية المركزية ، ومن ثم فإن توزيع المعاينة يتبع التوزيع المعتدل بصرف النظر عن التوزيع الأصلي للمجتمع . ومن ثم فإن مقدار فترة ثقة ١٠٠ (١ - α) ٪ لمتوسط

مجتمع معلوم التباين هو :

$$\bar{x} \pm \frac{\sigma}{\sqrt{n}} \frac{1}{\sqrt{\alpha}} Z_{\frac{\alpha}{2}} \\ 1000 \pm \frac{150}{100\sqrt{}} 1.96 \\ 1000 \pm 29.4$$

الحد الأدنى لفترة الثقة = ٩٧٠,٦ ساعة .

الحد الأعلى لفترة الثقة = ١٠٢٩,٤ ساعة .

وهذا يعني أنه بدرجة ثقة ٩٥ ٪ تحتوي الفترة من ٩٧٠,٦ ساعة إلى

١٠٢٩,٤ ساعة على الوسط الحسابي لعمر المصباح في المجتمع .

$$(ح) \quad \alpha - 1 = \alpha - 0,99 = 0,01, \quad \alpha \frac{1}{4} = 0,0005 = Z_{\dots} = 2,0705$$

ويكون تقدير فترة الثقة ٩٩ ٪ لمتوسط عمر المصباح هو :

$$\frac{100}{1.01\sqrt{}} 2,0705 \pm 1000 \\ 28,620 \pm 1000$$

الحد الأدنى لفترة الثقة = ٩٦١,٣٧٥ ساعة .

الحد الأعلى لفترة الثقة = ١٠٣٨,٦٢٥ ساعة .

وهذا يعني أنه بدرجة ثقة ٩٩ ٪ تحتوي الفترة من ٩٦١,٤ ساعة إلى

١٠٣٨,٦ ساعة على الوسط الحسابي لعمر المصباح في المجتمع .

(د) ن = ٢٠٠

ويكون تقدير فترة الثقة ٩٥ ٪ لمتوسط عمر المصباح هو :

$$\frac{100}{2.00\sqrt{}} 1,96 \pm 1000 \\ 20,789 \pm 1000$$

الحد الأدنى لفترة الثقة = ٩٧٩,٢١١ ساعة .

الحد الأعلى لفترة الثقة = ١٠٢٠,٧٨٩ ساعة .

وهذا يعني أنه بدرجة ثقة ٩٥ ٪ تحتوي الفترة من ٩٧٩,٢ ساعة إلى

١٠٢٠,٨ ساعة على الوسط الحسابي لعمر المصباح في المجتمع .

العوامل المحددة لفترة الثقة :

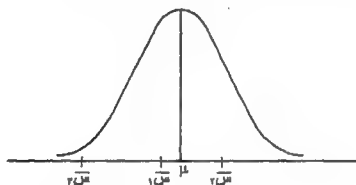
بالنظر إلى قانون (٨) نجد أن العوامل المحددة لفترة الثقة هي :

$Z_{\frac{\alpha}{2}}$ ، n ، σ . إلا أن الباحث لا يستطيع التحكم في الانحراف المعياري للمجتمع لكنه يستطيع التحكم في مستوى المعنوية وفي حجم العينة . ولقد رأينا في مثال (٣) : أنه عند درجة ثقة ٩٩ ٪ فإن فترة الثقة كانت (٩٦١,٤ ، ١٠٣٨,٦) وعند درجة ثقة ٩٥ ٪ فإن فترة الثقة كانت (٩٧٠,٦ ، ١٠٢٩,٤) وهذا يعني أنه إذا انخفضت درجة الثقة فإن فترة الثقة تصبح أضيق ، كما رأينا أيضاً أنه إذا ازداد حجم العينة من ١٠٠ مفردة إلى ٢٠٠ مفردة — مع الاحتفاظ بنفس درجة الثقة ٩٥ ٪ فإن فترة الثقة تصبح أضيق : من (٩٧٠,٦ ، ١٠٢٩,٤) إلى (٩٧٩,٢ ، ١٠٢٠,٨) . وعموماً يفضل الباحث فترة ثقة ضيقة لأن هذا يعطي نتيجة أكثر دقة بالنسبة لموقع معلمة المجتمع ، ويستطيع تحقيق ذلك إما بتخفيض درجة الثقة أو بزيادة حجم العينة . ولكن تخفيض درجة الثقة يخفض احتمال أن تحتوي فترة الثقة على معلمة المجتمع ، وهذا أمر غير مرغوب فيه . لذلك من الأفضل زيادة حجم العينة إذا ما أردنا تضيق فترة الثقة وبالتالي الحصول على نتائج أكثر دقة فيما يتعلق بموقع معلمة المجتمع .

تفسير درجة الثقة :

ما هو التفسير لدرجة ثقة ٩٥ ٪ مثلاً ؟ ففي المثال السابق إذا أخذنا جميع العينات العشوائية الممكنة التي حجمها n من هذا المجتمع ، وحسبنا في كل منها الأوساط الحسابية ، ثم وضعنا فترة ثقة ٩٥ ٪ للمعلمة μ حول كل وسط حسابي ، يمكننا التوقع بأن ٩٥ ٪ من هذه الفترات ستحتوي على المعلمة μ و ٥ ٪ منها لا تحتوي عليها . ويبين شكل (٥) الأوساط الحسابية \bar{x}_1 ، \bar{x}_2 و \bar{x}_3 لثلاث عينات عشوائية مسحوبة من نفس المجتمع ، كما يبين

الشكل فترات الثقة حول هذه المتوسطات . ومن الواضح في هذا الشكل أن فترات الثقة حول \bar{S}_1 ، \bar{S}_2 تحتوي بداخلها على μ . ويمكن القول بأنه بدرجة ثقة ٩٥ ٪ إذا أخذنا جميع العينات الممكنة التي لها نفس الحجم من هذا المجتمع ووضعنا فترة ثقة ٩٥ ٪ حول الأوساط الحسابية لهذه العينات ، فسيكون هناك ٩٥ ٪ من هذه الفترات مثل تلك الفترات التي حول \bar{S}_1 ، \bar{S}_2 والتي تحتوي بداخلها على μ ، كما سيكون هناك ٥ ٪ من هذه الفترات مثل تلك الفترة التي حول \bar{S}_3 فهي لا تحتوي بداخلها على μ .



$$\bar{S}_1 - 1.96\sigma_{\bar{S}_1} \quad \bar{S}_1 \quad \bar{S}_1 + 1.96\sigma_{\bar{S}_1}$$

$$\bar{S}_2 - 1.96\sigma_{\bar{S}_2} \quad \bar{S}_2 \quad \bar{S}_2 + 1.96\sigma_{\bar{S}_2}$$

$$\bar{S}_3 - 1.96\sigma_{\bar{S}_3} \quad \bar{S}_3 \quad \bar{S}_3 + 1.96\sigma_{\bar{S}_3}$$

شكل (٥) : التوزيع العيني للأوساط الحسابية \bar{S}
موضحا ثلاث فترات ثقة للمعلمة μ

ب - إذا كانت σ غير معلومة :

لقد افترضنا في دراستنا لفترات الثقة أن σ معلومة ، ولكن في كثير من الأحيان تكون σ غير معلومة ، ففي هذه الحالة نستخدم الانحراف المعياري للعينة كـ تقدير نقطة للمعلمة σ . وبما أن حجم العينة كبير ($n \geq 30$) ، فطبقاً لنظرية النهاية المركزية فإن التوزيع العيني للأوساط الحسابية يخضع تقريباً للتوزيع المعتدل وبالتالي فإن مقدار فترة الثقة في هذه الحالة يصبح كما يلي :

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \quad (9)$$

هو مقدار فترة ثقة $100(1 - \alpha)\%$ لمتوسط مجتمع مجهول التباين عندما $n \leq 30$.

مثال (٤) :

سحبت عينة عشوائية من أجور ٥٠ عامل من عمال أحد المصانع ، وفيما يلي التوزيع التكراري لأجور هؤلاء العمال :

فئات الأجر (بالجنيهات)	- ١٠٠	- ١٥٠	- ٢٠٠	- ٢٥٠	- ٣٠٠	٣٥٠ وأقل من ٤٠٠	المجموع
عدد العمال	٥	١٠	١٥	٨	٧	٥	٥٠

والمطلوب :

أ - إيجاد تقدير نقطة لمتوسط أجر العمال في هذا المصنع .

ب - إيجاد تقدير فترة ثقة ٩٩% لمتوسط أجر العمال في هذا المصنع .

الحل :

فئات الأجر (بالجنهات)	عدد العمال ك	مراكز الفئات س	م ك	س ^٢ ك
١٠٠ -	٥	١٢٥	٦٢٥	٧٨١٢٥
١٥٠ -	١٠	١٧٥	١٧٥٠	٣٠٦٢٥٠
٢٠٠ -	١٥	٢٢٥	٣٣٢٥	٧٥٩٣٧٥
٢٥٠ -	٨	٢٧٥	٢٢٠٠	٦٠٥٠٠٠
٣٠٠ -	٧	٣٢٥	٢٢٧٥	٧٣٩٣٧٥
٣٥٠ وأقل من ٤٠٠	٥	٣٧٥	١٨٧٥	٧٠٣١٢٥
المجموع	٥٠		١٢١٠٠	٣١٩١٢٥٠

١ - مقدر نقطة لمتوسط أجر العمال في هذا المصنع = $\bar{x} = \frac{\sum x_k}{n}$

$$\bar{x} = \frac{\sum x_k}{n} = \frac{12100}{50} = 242 \text{ جنيه}$$

٢ - بما أن حجم العينة كبير ($n = 50$) ، فطبقا لنظرية النهاية

المركزية فإن التوزيع العيني للأوساط الحسابية يخضع للتوزيع المعتدل ،
وحيث أن σ مجهولة فإننا نستخدم الانحراف المعياري للعينة كـ $\hat{\sigma}$ كمقدر
نقطة لمعلمة المجتمع σ . وحيث :^(١)

$$Z = \frac{\bar{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} = \frac{\bar{x} - \mu}{\frac{\sqrt{\frac{\sum x_k^2}{n} - \bar{x}^2}}{\sqrt{n}}}$$

$$= \frac{\bar{x} - \mu}{\sqrt{\frac{\sum x_k^2}{n} - \bar{x}^2}}$$

$$= \frac{242 - \mu}{\sqrt{\frac{3191250}{50} - 242^2}} = 5368.367$$

^(١) لقد قمنا بالقسمة على $n - 1$ (درجات الحرية) لأن الانحراف المعياري محسوب من عينة .
إلا أنه يمكن القسمة على n حيث أن العينة كبيرة .

$$\sqrt{5368,367} = \varepsilon$$

$$= 73,269 \text{ جنيهاً}$$

ويكون مقدر فترة ثقة ١٠٠ (١ - α) % هو :

$$\bar{M} \pm \frac{\varepsilon}{\sqrt{n}} \alpha \frac{1}{\gamma} Z$$

حيث : $\bar{M} = 242$ جنيه ، $\varepsilon = 73,269$ جنيها ، $1 - \alpha = 0,99$ ،

$$\alpha = 0,01 \quad ، \quad \alpha \frac{1}{\gamma} = 0,005 \quad ، \quad Z = 2,575$$

∴ تقدير فترة ثقة ٩٩ % للمعلمة μ هو :

$$242 \pm \frac{73,269}{\sqrt{50}}$$

$$242 \pm 10,372$$

الحد الأدنى لفترة الثقة = ٢٣١,٦٢٨ جنيها .

الحد الأعلى لفترة الثقة = ٢٥٢,٣٨٢ جنيها .

وهذا يعني أنه بدرجة ثقة ٩٩ % تحتوي الفترة من ٢٣١,٦٢٨ جنيه إلى ٢٥٢,٣٨٢ جنيه على الوسط الحسابي لأجر العمال في المجتمع .

(٢ - ٤) تقدير فترة ثقة لمتوسط مجتمع في حالة

العينات الصغيرة :

رأينا في المبحث السابق أنه عندما يكون حجم العينة كبيراً ($n \geq 30$) فإننا نستخدم التوزيع المعتدل سواء كانت σ معلومة أم غير معلومة . وذلك لأنه

وفقا لنظرية النهاية لمركزية ، عندما تكون العينات كبيرة فإن التوزيع العيني للأوساط الحسابية س يكون معتدلا تقريبا وذلك بصرف النظر عن شكل التوزيع الأصلي .

وفي كثير من الأحيان يتعذر الحصول على عينة كبيرة سواء بسبب تكلفتها الباهظة أو بسبب طبيعة التجربة نفسها ، فمثلا يزيد المستثمر معرفة ربح السهم قبل قيامه بعملية لل شراء مما يستلزم آراء كثير من بيوت الخبرة ، وما يتطلبه ذلك من تكاليف باهظة مما يجعل المستثمر يكتفي بعينة صغيرة ، مثال آخر هو اختبار دواء جديد للشفاء من مرض معين ، هنا أيضا يتم استخدام عينة صغيرة بسبب عدم وجود مرضى كثيرين مصابين بهذا المرض وعلى استعداد لتجربة الدواء الجديد .

فإذا كان حجم العينة صغيرا ($n > 30$) يجب التفرقة بين حالة ما إذا كانت σ معلومة أم لا .

أ - إذا كانت σ معلومة :

إذا كانت σ معلومة ، وكان التوزيع الأصلي للمجتمع الذي أخذت منه العينة هو توزيعا معتدلا ، ففي هذه الحالة يمكننا استخدام التوزيع المعتدل لتقدير فترة الثقة .

وفي هذه الحالة فإن :

$$(10) \quad \bar{x} \pm Z \frac{\sigma}{\sqrt{n}}$$

هو مقدر فترة ثقة $100(\alpha - 1)\%$ لمتوسط مجتمع معتدل ومعلوم التباين عندما $n > 30$.

مثال (٥) :

سحبت عينة عشوائية حجمها ٥ مفردات من مجتمع له توزيع معتدل
ع (μ ، σ) وكانت المشاهدات كما يلي :

$$٣٠ ، ٣٥ ، ٢٨ ، ٢٥ ، ٢٠$$

والمطلوب : (١) إيجاد فترة ثقة ٩٠ ٪ للوسط الحسابي للمجتمع .

(٢) إيجاد فترة ثقة ٩٥ ٪ للوسط الحسابي للمجتمع .

الحل :

$$(١) \quad \sigma = ٥ ، \quad \mu = ٢٠$$

$$١,٦٤٥ = Z_{\alpha/2} ، \quad \alpha/2 = ٠,٠٥ ، \quad \alpha = ٠,١٠ ، \quad \alpha - ١ = ٠,٩٠$$

بما أن المجتمع الذي سحبت منه العينة هو مجتمع معتدل ، σ معلومة ،
فإن التوزيع العيني للأوساط الحسابية يتبع التوزيع المعتدل ، ومن ثم :

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

هي مقدار فترة ثقة ١٠٠ ($\alpha - ١$) ٪ لمتوسط المجتمع μ .

$$\bar{X} = \frac{\text{مجموع}}{n} = \frac{٣٠ + ٣٥ + ٢٨ + ٢٥ + ٢٠}{٥}$$

$$= ٢٧,٦$$

∴ تقدير فترة ثقة ٩٠ ٪ للمعلمة μ هو :

$$\left(\frac{Y}{\sqrt{n}} \right) 1,640 \pm 27,6$$

$$1,471 \pm 27,6$$

الحد الأدنى لفترة الثقة = ٢٦,١٢٩

الحد الأعلى لفترة الثقة = ٢٩,٠٧١

وهذا يعني أنه بدرجة ثقة ٩٠ ٪ تتسوي الفترة من ٢٦,١٢٩ إلى ٢٩,٠٧١ على الوسط الحسابي للمجتمع μ .

$$(2) \quad 1 - \alpha = 0,90 \quad , \quad \alpha = 0,10 \quad , \quad \frac{1}{Y} = 0,25 = Z_{0,25} \quad , \quad 1,96$$

ويكون تقدير فترة ثقة ٩٠ ٪ لمتوسط المجتمع μ هو :

$$\left(\frac{Y}{\sqrt{n}} \right) 1,96 \pm 27,6$$

$$1,703 \pm 27,6$$

الحد الأدنى لفترة الثقة = ٢٥,٨٤٧

الحد الأعلى لفترة الثقة = ٢٩,٣٥٣

وهذا يعني أنه بدرجة ثقة ٩٠ ٪ تحتوي الفترة من ٢٥,٨٤٧ إلى ٢٩,٣٥٣ على الوسط الحسابي للمجتمع μ .

ب - إذا كانت σ غير معلومة :

وفي كثير من الأحيان تكون σ غير معلومة . فإذا حدث هذا وكان التوزيع الذي سحبت منه العينة توزيعاً معنوياً أو قريباً من الاعتدال ، نستخدم الانحراف المعياري s للعينة كمقدر نقطة للمعلمة σ . وفي هذه الحالة لا يمكننا استخدام التوزيع المعنوي في تقدير فترة ثقة للمعلمة μ . وهذا لأن المتغير العشوائي :

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = t$$

يتبع توزيع t بدرجات حرية $(n - 1)$.

وإذا كانت $t_{(n-1, 1-\frac{\alpha}{2})}$ هي قيمة t التي تجعل إلى يمينها مساحة قدرها $\frac{\alpha}{2}$ فإن :

(١١)

$$\bar{x} \pm t_{(n-1, 1-\frac{\alpha}{2})} \frac{s}{\sqrt{n}}$$

هو مقدر فترة ثقة $100(1 - \alpha)\%$ لمتوسط مجتمع معنوي أو قريب من الاعتدال وغير معلوم التباين عندما $n > 30$.

مثال (٦) :

أراد أحد المستثمرين تقدير متوسط العائد المتوقع للسهم الذي تصدره إحدى الشركات ، ولقد قام بالاستعانة بخمس بيوت الخبرة في سوق الأوراق المالية ، وكانت توقعاتها كالتالي (بالجنهات) :

١٠,٥ ، ١٦,٢ ، ١١,٤ ، ٩,٦

فإذا علمت أن مجتمع عائد السهم يتوزع توزيعاً قريباً من الاعتدال ، فالمطلوب إيجاد فترة ثقة ٩٥ ٪ لمتوسط العائد على السهم في هذه الشركة .

الحل :

بما أن التوزيع قريب من الاعتدال ، وغير معلوم للتباين والعينة صغيرة ($n = ٥$) فإن :

مقدر فترة ثقة ١٠٠ ($١ - \alpha$) ٪ لمتوسط المجتمع هو :

$$\bar{x} \pm t_{(n-1, 1-\frac{\alpha}{2})} \frac{s}{\sqrt{n}}$$

لذلك يجب أولاً حساب \bar{x} ، s .

س	س - \bar{x}	(س - \bar{x}) ^٢
١٠,٥	١,٩-	٣,٦١
١٦,٢	٣,٨	١٤,٤٤
١١,٤	١-	١
١٤,٣	١,٩	٣,٦١
٩,٦	٢,٨-	٧,٨٤
٦٢		٣٠,٥

$$\bar{x} = \frac{\sum x}{n} = \frac{62}{5} = 12,4 \text{ جنيهها}$$

$$s^2 = \frac{1}{n-1} \sum (x - \bar{x})^2$$

$$= \frac{1}{5-1} \sum (12,4 - x)^2$$

$$\frac{1}{4} = (30.5)$$

$$= 7,625$$

$$ع = 7,625 = 2,761 \text{ جنيها}$$

$$1 - \alpha = 0.95 , \alpha = 0.05 , \frac{1}{4} = \alpha = 0.25 , n = 5$$

وبالبحث في جدول t نجد أن :

$$t(0.25, 4) = 2,776$$

ومن ثم فإن تقدير فترة ثقة ٩٥ ٪ لمتوسط عائد للسهم هو :

$$2,776 \pm 12,4 \left(\frac{2,761}{\sqrt{5}} \right)$$

$$3,438 \pm 12,4$$

الحد الأدنى لفترة الثقة = ٢ ° ٨

الحد الأعلى لفترة الثقة = ١٥,٨٢٨

وهذا يعني أنه بدرجة ثقة ٩٥ ٪ غوي الفترة من ٨,٩٧٢ جنيها إلى

١٥,٨٢٨ على عائد السهم الحقيقي في هذه الشركة .

(٢ - ٥) تقدير فترة ثقة للنسبة في المجتمع :

في كثير من الأحيان يتم تقسيم المجتمع إلى نوعين من المفردات :

المفردات التي تنصف بصفة معينة والمفردات التي لا تنصف بهذه الصفة .

فمثلا قد ينقسم المجتمع إلى مدخن وغير مدخن ، أو إلى إناث وذكور ، وقد

ينقسم مجتمع إنتاج مصنع معين إلى إنتاج معيب وإنتاج غير معيب . فإذا

سحبت مفردة من أحد هذه المجتمعات بطريقة عشوائية ، فإن احتمال تمتع هذه المفردة بصفة معينة = نسبة هذه الظاهرة في المجتمع = θ . فمثلاً إذا كانت نسبة المدخنين في مجتمع معين : $\theta = 0,40$ ، فإن نسبة غير المدخنين في هذا المجتمع ، $(1 - \theta) = 1 - 0,40 = 0,60$.

وإذا سحبنا عينة عشوائية حجمها n من هذا المجتمع ، وإذا كان s هو المتغير العشوائي الذي يمثل عدد الذين يتصفون بهذه الصفة في العينة ، فإن s تكون متغيراً عشوائياً له توزيع ذي الحدين .

وفي كثير من الأحيان تكون النسبة في المجتمع (θ) مجهولة ، ونريد تقدير هذه النسبة بنقطة أو بفترة .

أ - التقدير بنقطة :

يمكن استخدام النسبة في العينة كمقدر نقطة لنسبة المجتمع θ . فإذا استخدمنا الرمز q للدلالة على النسبة في العينة حيث :

$$q = \frac{\text{عدد المفردات التي لها الصفة}}{\text{حجم العينة}} = \frac{s}{n}$$

ويكون مقدر نقطة لنسبة المجتمع θ هو :

(١٢)

$$\hat{\theta} = q = \frac{s}{n}$$

ب - التقدير بفترة :

$\hat{\theta}$ متغير عشوائي له توزيع احتمالي معين . وسنكتفي في دراستنا هنا بأخذ حالة كون حجم العينة كبيراً بحيث تكون : $0 \leq \theta \leq 1$ ، $0 \leq \hat{\theta} \leq 1$. وعندما تكون θ غير معلومة ، نستخدم بدلاً من θ مقدر النقطة لها $\hat{\theta}$ ،

ومن ثم لكي تكون العينة كبيرة يجب أن تكون $n \leq \theta$ ، $n \leq (\theta - 1)$ ومن ثم لا نستطيع حساب تباين التوزيع العيني $\hat{\theta}$ من التوزيع المعتدل الذي متوسطه : $\theta - \hat{\theta}$

$$\frac{(\theta - 1)\theta}{n} = \sigma^2_{\theta}$$

وبما أننا نريد تقدير θ ، إذن θ غير معلومة ، ومن ثم لا نستطيع حساب تباين التوزيع العيني σ^2_{θ} ، لذلك نستخدم تباين العينة $\sigma^2_{\hat{\theta}}$ كمقدر نقطة لتباين التوزيع العيني σ^2_{θ} حيث :

$$\frac{(\hat{\theta} - 1)\hat{\theta}}{n} = \sigma^2_{\hat{\theta}}$$

ومن ثم فإن :

$$(13) \quad \frac{(\hat{\theta} - 1)\hat{\theta}}{n} \sqrt{\alpha \pm Z \pm \hat{\theta}}$$

هو مقدر فترة ثقة $100(\alpha - 1)\%$ لنسبة مجتمع في حالة العينات الكبيرة حيث : $n \leq \hat{\theta}$ ، $n \leq (\hat{\theta} - 1)$

مثال (٧) :

تريد أحد الشركات القيام بتسويق نوع جديد من مسحوق الغسيل . وقبل القيام بذلك أرادت الشركة القيام بأبحاث للسوق لمعرفة مدى تفضيل الناس لهذا المسحوق . فسحبت عينة عشوائية من ٢٠٠ مستهلك وأهدت لهم عبوة مجانية ، وبعد استعمالها وجدت الشركة أن ١٤٠ منهم فضلوا هذا المسحوق . المطلوب :

أ - تقدير نقطة لنسبة المستهلكين الذين يفضلون هذا المسحوق .

ب - تقدير فترة ثقة ٩٥ % لنسبة المستهلكين الذين يفضلون هذا المسحوق .

الحل :

$$أ - ن = ٢٠٠ ، م = ١٤٠$$

$$\hat{\theta} = ق = \frac{١٤٠}{٢٠٠} = ٠,٧ \text{ هي مقدار نقطة لنسبة المجتمع } \theta$$

$$ب - ١ - \alpha = ٠,٩٥ ، \alpha = ٠,٠٥ ، \frac{1}{\sqrt{n}} = ٠,٠٢٥ ، Z = ١,٩٦$$

$$ن \hat{\theta} = (٠,٧) ٢٠٠ = ١٤٠ < ٥$$

$$ن (\hat{\theta} - ١) = (٠,٧ - ١) ٢٠٠ = -٦٠ < ٥$$

$$ن \hat{\theta} < ٥ ، ن (\hat{\theta} - ١) < ٥ ، إذن العينة كبيرة ،$$

و بتطبيق نظرية النهاية المركزية نجد أن :

$$\frac{(\hat{\theta} - ١) \hat{\theta}}{ن} \sqrt{n} \sim Z \pm \hat{\theta}$$

هو مقدار فترة ثقة ١٠٠ (١ - α) % لنسبة مجتمع .

أي أن :

$$\frac{(٠,٣) ٠,٧}{٢٠٠} \sqrt{n} \sim ١,٩٦ \pm ٠,٧$$

$$(٠,٠٣٢٤) ١,٩٦ \pm ٠,٧$$

$$٠,٠٦٣٥ \pm ٠,٧$$

الحد الأدنى لفترة الثقة = ٠,٦٣٦٥

الحد الأعلى لفترة الثقة = ٠,٧٣٦٥

أي أنه بدرجة ثقة ٩٥ ٪ تحتوي الفترة من ٠,٦٣٦٥ إلى ٠,٧٦٣٥ على النسبة الحقيقية لتفضيل المستهلكين لهذا المسحوق .

(٢ - ٦) تحديد العينة لتقدير متوسط مجتمع :

في الأمثلة السابقة كان حجم العينة معرّماً : ولم نتعرض لسبب اختيار حجم العينة . ولنفرض أننا نريد تقدير بفترة لمتوسط مجتمع ، فما سر حجم العينة الواجب سحبها ؟ فعلى سبيل المثال إذا كان من الممكن الحصول على فترة الثقة التي نريدها من عينة حجمها ٥٠ مفردة ، فعند أخذ عينة حجمها ٢٠٠ مفردة نكون قد أضعنا كثير من النفقات والوقت والجهد بدون مبرر . فإذا كنا نعلم مستوى الثقة وطول فترة الثقة التي نريدها ، يمكننا معرفة حجم العينة التي تعطينا هذه النتائج .

وكما رأينا سلفاً فإن مقدار فترة ثقة $1 - \alpha$ ٪ لمتوسط مجتمع هو :

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

وبافتراض أن سحب العينة يتم بإرجاع أو أن تكون النسبة بين حجم العينة n وحجم المجتمع N هي : $\frac{n}{N} > 0.05$ ، فإن :

$$Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \text{ يسمى بخطأ التقدير } (E)$$

$$(14) \quad Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = E \quad \text{أي أن}$$

ونحصل على قيمة n بحل معادلة (١٤) نجد أن :

$$(15) \quad n = \left[\frac{Z_{\frac{\alpha}{2}} \sigma}{E} \right]^2$$

وهنا σ مجهولة ، ويمكن تقديرها باستخدام الانحراف المعياري ع المحسوب من عينة مبدئية ذات حجم صغير متفق عليه ، أو يمكن تقديرها من عينة سبق الحصول عليها من دراسات سابقة أو دراسات مماثلة . وإذا كان توزيع المجتمع معتدلا تقريبا فإن :

$$\sigma \approx \sigma_6 = \text{المدى}$$

$$(16) \quad \text{ومن ثم } \sigma \approx \frac{\text{المدى}}{6} = \frac{\text{أكبر قيمة} - \text{أصغر قيمة}}{6}$$

ولكن إذا كان سحب العينة يتم بدون إرجاع ، أو عندما تكون النسبة بين حجم العينة n وحجم المجتمع M هي : $\frac{n}{M} \leq 0.05$ ، فيجب استخدام معامل التصحيح correction factor كما سبق وبيننا في الفصل الأول

$$(17) \quad \text{حيث : معامل التصحيح} = \frac{n - \mu}{1 - \mu}$$

ويصبح خطأ التقدير في هذه الحالة :

$$(18) \quad Z = E \cdot \alpha \cdot \frac{\sigma}{\sqrt{\frac{n - \mu}{1 - \mu}}}$$

وبحل هذه المعادلة نحصل على قيمة n ،

$$(19) \quad n = \frac{\sigma^2 \alpha^2 Z^2 \mu}{\sigma^2 \alpha^2 Z^2 + (1 - \mu) E^2}$$

مثال (٨) :

قام قسم البحوث في أحد الشركات بتقدير متوسط الوقت الذي يقضيه العاملين في هذه الشركة للوصول من منازلهم إلى مقر عملهم . فإذا علمت أن قسم البحوث يريد أن يكون التقدير صحيحا في حدود دقيقة واحدة من المتوسط

الأصلي وبدرجة ثقة ٩٩ ٪ ، وأنه بسحب عينة صغيرة مبدئية وجد أن الانحراف المعياري يساوي ٥ دقائق ، فالمطلوب : تحديد حجم العينة الواجب سحبها .
الحل :

$$E = 1 \text{ دقيقة} , \alpha - 1 = 0.99 , \alpha = 0.01 , \\ \frac{1}{\gamma} = \alpha = 0.005 , Z_{\dots} = 2.575 , \hat{\sigma} = \sigma = 5 \text{ دقائق} .$$

ويكون حجم العينة :

$$n = \left[\frac{\hat{\sigma}^2 \alpha \frac{1}{\gamma} Z^2}{E^2} \right] = \\ = \left[\frac{(5)^2 2.575^2}{1} \right] = 165.8 \approx 166 \text{ عامل}$$

وهذا يعني أنه يجب سحب عينة من ١٦٦ عامل لتقدير متوسط الوقت المنقضي لوصول العاملين من منازلهم إلى مقر عملهم ، وذلك حتى يكون التقدير صحيحا في حدود دقيقة واحدة من المتوسط الأصلي وبدرجة ثقة ٩٩ ٪ .

مثال (٩) :

في المثال السابق ، بالإضافة إلى المعلومات السابقة . إذا علمت أن عدد العاملين في هذه الشركة هو ٢٥٠٠ عامل ، المطلوب تحديد حجم العينة الواجب سحبها .

الحل :

بالإضافة إلى المعلومات السابقة فإن حجم المجتمع م = ٢٥٠٠ عامل

$$\begin{aligned}
 \frac{{}^2\sigma_{\alpha} + \frac{{}^2Z}{\sqrt{}}}{\sqrt{{}^2\sigma_{\alpha} + \frac{{}^2Z}{\sqrt{}} + (1 - \rho) {}^2E}} &= \text{ن} \\
 \frac{{}^2(5) {}^2(2,075) 2,000}{\sqrt{{}^2(5) {}^2(2,075) + (1 - 2000) {}^21}} &= \\
 \frac{(160,766) 2,000}{160,766 + 2499} &= \\
 \frac{414414,063}{2664,766} &= \\
 156 \approx 100,0 &= \text{عامل}
 \end{aligned}$$

(٢ - ٧) تحديد حجم العينة لتقدير نسبة المجتمع :

وكما فعلنا بالنسبة لتحديد حجم العينة لتقدير متوسط مجتمع ، سنحدد في هذا المبحث حجم العينة لتقدير نسبة المجتمع . فإذا كانت نسبة حجم العينة ن إلى حجم المجتمع م هي $\frac{ن}{م} > 0,05$ ، فإن خطأ التقدير هو :

$$(٢٠) \quad \frac{(\theta - 1) \theta}{\text{ن}} \sqrt{{}^2\sigma_{\alpha} + \frac{{}^2Z}{\sqrt{}}} = E$$

ومنها نجد أن :

$$(٢١) \quad \frac{(\theta - 1) \theta {}^2\sigma_{\alpha} + \frac{{}^2Z}{\sqrt{}}}{{}^2E} = \text{ن}$$

أما إذا كانت نسبة حجم العينة n إلى حجم المجتمع M هي : $\frac{n}{M} \leq 0,05$ ، فإن :

$$(22) \quad n = \frac{(\theta - 1) \theta_{\alpha} \frac{1}{\sqrt{Y}} Z_{\frac{1}{2}}}{(\theta - 1) \theta_{\alpha} \frac{1}{\sqrt{Y}} Z + (1 - \theta) \sqrt{E}}$$

وفي جميع الأحوال استخدمنا هنا θ معلمة المجتمع المجهولة المراد تقديرها . وكما فعلنا في المبحث السابق يمكننا أخذ عينة مبدئية صغيرة ومنها حساب مقتر نقطة $\hat{\theta}$ للنسبة θ . أو يمكننا استخدام $\theta = 0,5$ والتعويض عنها في معادلة (٢١) أو (٢٢) ، ولكن هذه الطريقة تجعل حجم العينة أكبر مما يمكن ، هذا لأن ضرب $0,5$ في $0,5$ يعطي مقداراً أكبر من ضرب أي نسبتيين أخريتين لكل من θ ، $(1 - \theta)$.

مثال (١٠) :

يتعهد أحد المطاعم بتوصيل الطلبات إلى المنازل خلال ٣٠ دقيقة من طلب الطلبية . وأرادت إدارة هذا المطعم تقدير نسبة الطلبيات التي وصلت إلى المستهلكين خلال ٣٠ دقيقة . فما هو حجم العينة اللولجب أخذه حتى يصبح خطأ المعاينة $0,02$ من نسبة المجتمع ، بدرجة ثقة 99% .

الحل :

$$E = 0,02 \quad , \quad \alpha = 0,99 \quad , \quad \alpha - 1 = 0,01$$

$$\frac{1}{Y} = \alpha = 0,005 \quad , \quad Z_{\frac{1}{2}} = 2,575$$

إذن فترة الثقة 99% هي : $\hat{\theta} \pm 0,02$

وسنأخذ كتقدير لنسبة المجتمع النسبة $\hat{\theta} = 0,0$ ، ومن ثم فإن حجم العينة سيكون :

$$n = \frac{(\theta - 1) \theta + \frac{1}{\gamma} Z^2}{\gamma E}$$

$$4144 \text{ مستهلك} = \frac{1,658}{0,0004} = \frac{(0,0)(0,0) + (2,575)^2}{\gamma(0,02)}$$

وهذا يعني أن على مدير المطعم أخذ عينة من 4144 مستهلك .

تمارين (٢)

١ - في إحدى الشركات تم تدريب عينة عشوائية من ٦٠ موظف على إنجاز عمل معين . ولقد قام فريق من الباحثين بقياس الزمن الذي يستغرقه كل موظف في إنجاز هذا العمل ، فكان متوسط هذا الزمن ١٥ دقيقة بانحراف معياري ٣ دقائق .

والمطلوب :

- أ - إيجاد تقدير نقطة لمتوسط الزمن المستغرق في هذا العمل .
ب - إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط الزمن الذي يتخذه الموظف في إنجاز هذا العمل .

٢ - تقوم إحدى الشركات بتعبئة العكر في أكياس من البلاستيك . ولمعرفة متوسط وزن الكيس قامت الشركة بسحب عينة عشوائية من ٥٠ كيس فوجدت أن متوسط وزن الكيس ٩٠٠ جرام بانحراف معياري ٢٥ جرام .
والمطلوب :

- أ - إيجاد تقدير نقطة لمتوسط وزن الكيس .
ب - إيجاد تقدير فترة ثقة ٩٠ ٪ لمتوسط وزن الكيس .

٣ - قام قسم البحوث بإحدى شركات الطيران بعمل دراسة لمعرفة عدد المقاعد الشاغرة على رحلات طيرانها . فصحبت عينة عشوائية من ١٠٠ رحلة طيران ، فوجد أن الوسط الحسابي للأماكن الشاغرة بها هي : ١٥,٢ مقعد ، بانحراف معياري ٥,٢ مقعد .

والمطلوب :

- أ - إيجاد تقدير نقطة لمتوسط المقاعد الشاغرة .
ب - إيجاد تقدير فترة ثقة ٩٩ ٪ لمتوسط الأماكن الشاغرة .

٤ - أرادت مصلحة البريد معرفة متوسط أرصدة دفاتر البريد في أحد فروعها . فسحبت عينة عشوائية من ٢٥ دفتر توفير فوجدت أن الوسط الحسابي للأرصدة هو ٣٥٠٠ جنيه بانحراف معياري ٤٠٠ جنيه . والمطلوب :

- أ - إيجاد تقدير نقطة لمتوسط أرصدة دفاتر البريد في هذا الفرع .
ب - إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط أرصدة دفاتر البريد في هذا الفرع علما بأن أرصدة دفاتر البريد تتوزع توزيعا معتمدا .

٥ - أراد أحد المحلات التجارية الكبيرة معرفة متوسط ما ينفقه العملاء أثناء التسوق في المحل ، لذلك سحبت عينة عشوائية من ٩ عملاء ، فوجد أن ما أنفقه هؤلاء العملاء أثناء التسوق في المحل هي (بالجنيهات) :

١٢٠ ، ١٥٠ ، ١١٥ ، ١٦٥ ، ٧٥ ، ١٨٠ ، ٦٥ ، ١٤٣ ، ٩٧

والمطلوب :

- أ - إيجاد تقدير نقطة لكل من متوسط وتباين ما ينفقه العملاء أثناء التسوق في هذا المحل :

ب - إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط ما ينفقه العملاء أثناء التسوق في هذا المحل ، علما بأن المجتمع الذي سحبت منه العينة يتوزع توزيعا معتمدا .

٦ - أراد أحد مديري إنتاج أحد المصانع معرفة أقطار الكرات التي تنتجها أحد الآلات . فقام بسحب عينة عشوائية من ١٠ كرات فوجد أن متوسط قطر الكرات هو ٨٠,٥ ملليمتر بانحراف معياري

٥,٦ ملليمتر . والمطلوب : إيجاد تقدير فترة ثقة ٩٩ ٪ لمتوسط قطر الكرات من إنتاج هذه الآلة علما بأن المجتمع الذي سحبت منه العينة يتوزع توزيعاً معتدلاً .

٧ — لتقدير متوسط عدد حالات الطوارئ التي تصل خلال اليوم الواحد لإحدى المستشفيات ، قام مدير هذه المستشفى بسحب عينة عشوائية من ٢٥ يوما ، فوجد أن للوسط الحسابي = ٢١,٢ حالة والانحراف المعياري ٥,٤ حالة . والمطلوب : إيجاد تقدير فترة ثقة ٩٠ ٪ لمتوسط عدد الحالات التي تصل إلى قسم الطوارئ خلال اليوم الواحد ، علما بأن المجتمع الذي سحبت منه العينة يتوزع توزيعاً معتدلاً .

٨ — أراد أحد مصانع الساعات دراسة دقة نوع معين من الساعات التي ينتجها . فسحب عينة عشوائية من ٧ ساعات من إنتاج المصنع وقام برصد الزمن قبل وبعد ٤٨ ساعة ، فوجد أن عدد الثواني التي قدمتها أو أخرتها الساعة هي على التوالي :

$$+٣ ، -٥ ، +١٠ ، -٤ ، +٧ ، -٦ ، +٩$$

والمطلوب :

أ — إيجاد تقدير نقطة لكل من متوسط وتباين الزمن الذي تؤخره أو تقدمه هذه الساعات .

ب — إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط الزمن الذي تؤخره أو تقدمه هذه الساعات علما بأن المجتمع الذي سحبت منه العينة يتوزع توزيعاً معتدلاً .

٩ - لمعرفة نسبة الأمية الثقافية في الجامعة ، محبت عينة عشوائية من ١٥٠ طالب فوجد أن عدد الأميين ثقافيا هو ٦٥ طالب . فالمطلوب :

أ - إيجاد تقدير نقطة لنسبة الأمية الثقافية في الجامعة .

ب - إيجاد تقدير فترة ثقة ٩٥ ٪ لنسبة الأمية الثقافية في الجامعة

١٠ - قام أحد مراجعي الحسابات بمراجعة حسابات إحدى الشركات ، لذلك

قام بسحب عينة عشوائية من ٢٠٠ مستند فوجد فيها ٢٥ مستندا به أخطاء ..

والمطلوب :

أ - إيجاد تقدير نقطة لنسبة المستندات التي تحتوي على أخطاء .

ب - إيجاد تقدير فترة ثقة ٩٩ ٪ لنسبة المستندات التي تحتوي على أخطاء .

١١ - تريد إحدى الشركات القيام بعمل دراسة عن متوسط عدد أيام

الإجازات المرضية للموظفين بالشركة . فما هو عدد الموظفين الواجب

أخذه كعينة لإجراء هذه الدراسة علما بأن الشركة تريد أن يكون التقدير

صحيحا في حدود ثلاثة أيام من المتوسط الأصلي وبدرجة ثقة ٩٥ ٪ .

ولقد سبق لهذه الشركة أن قامت بدراسات سابقة استنتجت منها

أن الانحراف المعياري لعدد أيام الإجازات هو ٩ أيام .

١٢ - أرادت إحدى الشركات معرفة نسبة المستهلكين الذين يفضلون نسوع

الصابون الجديد الذي طرحته في الأسواق . ولقد قامت الشركة بدراسة

سابة على عينة صغيرة لمعرفة هذه النسبة فوجدت أنها ٠,٣٠ .

والمطلوب معرفة حجم العينة الواجب أخذه حتى يصبح الخطأ المسموح

به ٠,٠١ ، بدرجة ثقة ٩٩ ٪ .

الفصل الثالث

اختبارات الفروض الإحصائية

مقدمة :

لقد رأينا في الفصل السابق كيفية تقدير معلمية المجتمع المجهولة باستخدام إحصائية محسوبة من عينة عشوائية مسحوبة من هذا المجتمع . كما بينا أن هناك طريقتان لهذا التقدير : التقدير بنقطة والتقدير بفترة . وبالنسبة لمستوى معنوية معين ، كلما كانت فترة الثقة أضيق كلما زاد اعتقادنا في أن الإحصائية تمثل تقديراً دقيقاً لمعلمة المجتمع . ومن ناحية أخرى فعندما تكون العينة كبيرة فإن تطبيق نظرية النهاية المركزية تجعلنا لا نكتث بشكل توزيع المجتمع الأصلي ، طالما أن توزيع المعاينة يقترب من التوزيع المعتدل ، ومن ثم يمكننا استخدام القيمة المعيارية Z .

ولكن في كثير من الأحيان نجد أن البعض يدعي أن معلمة المجتمع تساوي قيمة معينة ، فمثلاً قد يدعي مدير إنتاج مصنع المصابيح الكهربائية أن متوسط عمر المصابيح في هذا المصنع هو $\mu = 1000$ ساعة . فإذا أخذنا عينة من إنتاج مصابيح هذا المصنع هل يمكننا إثبات ذلك ؟ بالطبع لا ، لأن الطريقة الوحيدة لمعرفة قيمة μ بدقة تتم عن طريق أخذ بيانات عن المجتمع بأسره أي كل إنتاج هذا المصنع في شهر معين مثلاً . ولكن هذه العينة تمكننا من قبول أو رفض إدعاء أن متوسط عمر المصابيح في هذا المصنع هو 1000 ساعة . وبما أن العينة ما هي إلا مجموعة من مفردات المجتمع ، لذلك فإن هذا الاستنتاج قد يكون خاطئاً . لإيضاح ذلك كله يقوم هذا الفصل بدراسة اختبارات

الفروض الإحصائية وذلك في أربعة مباحث . يتناول المبحث الأول شرح اختبارات الفروض الإحصائية ، ويتناول المبحث الثاني دراسة اختبارات الفروض الإحصائية المتعلقة بمتوسط مجتمع في حالة العينات الكبيرة ، ويتناول المبحث الثالث دراسة اختبارات الفروض الإحصائية المتعلقة بمتوسط مجتمع في حالة العينات الصغيرة ، ويتناول المبحث الرابع دراسة الفروض الإحصائية المتعلقة بنسبة مجتمع في حالة العينات الكبيرة .

(٣ - ١) الفروض الإحصائية :

سبق وذكرنا أنه لمعرفة معلمة المجتمع بدقة يجب إجراء الحصر الشامل . وإذا أخذنا مثال المصابيح الكهربائية ، فإن إدارة إنتاج المصنع أعلنت أن متوسط عمر المصباح من إنتاج هذا المصنع هو $\mu = 1000$ ساعة . ولنفرض أننا أخذنا عينة عشوائية من 100 مصباح من إنتاج هذا المصنع ووجدنا أن متوسط عمر المصباح في هذه العينة هو $\bar{x} = 970$ ساعة . فهل هذا يجعلنا نعتقد أن إدارة إنتاج هذا المصنع تدعي ادعاء خاطئاً لإيهام المستهلكين بأن عمر المصباح أكبر من عمره الحقيقي ، وأنه في الواقع عمر المصباح أقل من 1000 ساعة ؟ لا نستطيع اتهام مصنع المصابيح بهذا الاتهام إلا بعد أن تجري اختباراً للفروض الإحصائية ، لأن هذا الاتهام مبني على معلومات متخذة من عينة عشوائية . وقد يكون الفرق بين متوسط المجتمع ومتوسط العينة ناتج عن خطأ المعاينة فقط ، بمعنى أنه إذا سحبنا عينة عشوائية أخرى من نفس المجتمع فقد نجد أن متوسط عمر المصباح في العينة $\bar{x} = 1010$ ساعة مثلاً . لذلك فإننا نقوم بإجراء اختباراً للفروض الإحصائية لمعرفة ما إذا كان الفرق بين متوسط المجتمع $\mu = 1000$ ساعة ومتوسط العينة $\bar{x} = 970$ ساعة ناتجاً عن الصدفة فقط أم هو فرقاً حقيقياً

وفي الواقع فإن اختبار الفروض الإحصائية يشبه إلى حد كبير الاختبارات العلمية . فالعالم يقوم بوضع صياغة لنظرية معينة ثم بعد ذلك يقوم باختبار هذه النظرية عن طريق المشاهدات . وفي اختبارات الفروض الإحصائية فإن القائم بالبحث الإحصائي يقوم بوضع فرض معين بالنسبة لمعلمة المجتمع ، فهو يفترض أن معلمة المجتمع تساوي قيمة نظرية معينة . ثم بعد ذلك يقوم الباحث بسحب عينة عشوائية من هذا المجتمع ويقوم بمقارنة المشاهدات الناتجة من العينة بالافتراض النظري الذي وضعه : فإذا كانت المشاهدات لا تتفق مع الافتراض النظري فهو يرفض هذا الافتراض ، أما إذا كانت المشاهدات تتفق مع هذا الافتراض ، فإنه يقبله .

وبوجه عام فإن اختبارات الفروض تتضمن أربعة مراحل أساسية وهي :

١ - صياغة الفروض الإحصائية .

٢ - تعيين إحصائية الاختبار وحسابها .

٣ - تحديد مستوى المعنوية والمنطقة الحرجة .

٤ - اتخاذ القرار الإحصائي .

وفيما يلي سنقوم بدراسة كل من هذه المراحل على حدة .

١ - صياغة الفروض الإحصائية :

لو أخذنا مثال المصابيح الكهربائية فإن افتراض أن ما قاله مدير الإنتاج صحيحاً بأن متوسط عمر المصابيح من إنتاج هذا المصنع يساوي ١٠٠٠ ساعة ، يسمى بفرض العدم Null hypothesis ويرمز له بالرمز H_0 . ويكتب فرض العدم كما يلي :

$$H_0 : \mu = 1000 \text{ ساعة .}$$

وينص فرض العدم على أن القيمة النظرية لمعلمة المجتمع صحيحة إلى أن يثبت العكس . وبمعنى آخر ، فإن فرض العدم ينص على " عدم وجود فرق " بين معلمة المجتمع وإحصائية العينة ، ومن هنا جاءت تسمية فرض العدم . وفي مثالنا هذا يمكننا كتابة فرض العدم على الصورة الآتية :

$$H_0 : \mu \leq 1000 \text{ ساعة} .$$

وهذا لأن عمر المصابيح إذا زاد عن ١٠٠٠ ساعة فهذا أمر مرغوب فيه من وجهة نظر المستهلك ، لذلك فإن كتابة علامة $=$ أو \leq في فرض العدم لن تؤثر على الاختبار .

ويعتبر هذا المصنع غشاشاً من وجهة نظر المستهلك إذا كان متوسط عمر المصابيح أقل من ١٠٠٠ ساعة .

لذلك فإن الفرض الثاني يسمى بالفرض البديل ويرمز له بالرمز H_1 ويكتب على الصورة التالية :

$$H_1 : \mu > 1000 \text{ ساعة} ;$$

ويمكن القول بأن الفرض البديل هو الفرض الذي يكون صحيحاً إذا كان فرض العدم غير صحيح .

وفي مثالنا هذا يسمى هذا الاختبار اختبار طرف أيسر $a \text{ left-tailed test}$ لأن الفرض البديل هنا يحتوي على علامة $(>)$.

ولنأخذ مثالا آخر ، إذا كانت إحدى الشركات تباع آلة معينة لإنتاج إحدى قطع الغيار ، ولقد حددت الشركة بأن عدد الوحدات للعينة من إنتاج هذه الآلة في الشهر هو ٩٠ وحدة . هنا يكون فرض العدم :

$$H_0 : \mu = 90 .$$

وبما أنه من المستحب أن تكون عدد الوحدات المعيبة أقل من ٩٠ وحدة ، فيمكننا كتابة فرض العدم كما يلي :

$$H_0: \mu \geq 90$$

وبما أنه من غير المستحب (أو غير المرغوب فيه) أن تكون عدد الوحدات المعيبة أكبر من ٩٠ وحدة ، فإن الفرض البديل هو :

$$H_1: \mu < 90 \text{ وحدة}$$

ويسمى هذا الاختبار اختباراً طرَف أيمن $a \text{ right-tailed test}$ لأن الفرض البديل يحتوي على علامة (<)

ولنأخذ مثلاً آخر ، ولنفرض أن أحد المصانع ينتج نوع معين من المسامير قطره ٢ ملليمتر . فيكون فرض العدم هو :

$$H_0: \mu = 2 \text{ ملليمتر}$$

ومن المستحب في هذه الحالة أن يكون قطر هذه المسامير تساوي ٢ ملليمتر بالضبط ، ومن غير المستحب (أو المرغوب فيه) أن يزيد هذا القطر أو يقل عن ٢ ملليمتر ، لذلك فإن الفرض البديل هو :

$$H_1: \mu \neq 2 \text{ ملليمتر}$$

ويسمى مثل هذا الاختبار باختبار الطرفين $a \text{ two-tailed test}$ لأن الفرض البديل يحتوي على علامة (≠) .

وبصفة عامة فإن فرض العدم هو للفرض الذي ينص على أن معلومة المجتمع المعطاة صحيحة ، لذلك يحتوي فرض العدم دائماً على علامة (=) ، وقد يحتوي على إشارة (≥) أو (≤) . أما الفرض البديل فهو الفرض الذي ينص على غير المستحب بالنسبة لمعلومة المجتمع ، لذلك فهو لا يحتوي أبداً على علامة (=) ولكنه يحتوي على علامة (≠) أو (<) أو (>) .

٢ - تعيين إحصائية الاختبار وحسابها :

إحصائية الاختبار test statistic هي الإحصائية المستخدمة في اختبارات الفروض الإحصائية . وتتوقف قيمتها على بيانات العينة المسجوبة من المجتمع . ويمكن تعريف إحصائية الاختبار بأنها القاعدة أو المعيار المستخدم لاتخاذ القرار برفض أو قبول فرض العدم . وأغلب إحصائيات الاختبار تكون على الصورة :

$$\text{إحصائية الاختبار} = \frac{\text{إحصائية العينة} - \text{القيمة النظرية تلمعة}}{\text{الخطأ المعياري للتوزيع العيني}}$$

فعلى سبيل المثال إذا أردنا اختبار متوسط مجتمع ، وكسان التوزيع العيني يخضع للتوزيع المعتدل ، فإن قيمة Z المعيارية تعتبر إحصائية الاختبار في هذه الحالة ، حيث :

(١)

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

أما إذا كان التوزيع العيني يخضع لتوزيع t ، فإن قيمة t تعتبر إحصائية الاختبار في هذه الحالة ، حيث :

(٢)

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

٣ - تحديد مستوى المعنوية والمنطقة الحرجة :

في نهاية الاختبار يكون لدينا احتمالين : الأول رفض فرض العدم H_0 ، والثاني عدم رفض H_0 . والفرض غير المرفوض قد يكون صحيحاً أو غير صحيح ، كما أن الفرض المرفوض قد يكون صحيحاً أو غير صحيح . ومن ثم فإن لدينا أربع حالات ممكنة عند إجراء أي اختبار إحصائي ألا وهي :

١ - رفض فرض العدم بينما هو غير صحيح .

٢ - رفض فرض العدم بينما هو صحيح .

٣ - عدم رفض فرض العدم بينما هو غير صحيح .

٤ - عدم رفض فرض العدم بينما هو صحيح .

وتعتبر الحالتين (١) ، (٤) مرغوب فيها ، بينما تعتبر الحالات (٢) ، (٣) غير مرغوب فيها وتسمى بالأخطاء . وتسمى حالة رفض فرض العدم بينما هو صحيح بالخطأ من النوع الأول Type I error أو خطأ α error . ففي مثال المصابيح الكهربائية يحدث الخطأ من النوع الأول إذا كان متوسط عمر المصباح في الواقع يساوي ١٠٠٠ ساعة ، ولكن بالصدفة سحبنا عينة عشوائية من هذا المجتمع وكان متوسط عمر المصباح فيها أقل من ١٠٠٠ ساعة ، فقمنا باتخاذ قرار خاطئ برفض فرض العدم H_0 . وتسمى α بمستوى المعنوية Significance level وهي تمثل احتمال حدوث الخطأ من النوع الأول . وبوجه عام يحدد مستوى المعنوية α قبل القيام بالاختبار الإحصائي ، وعادة لا تتعدى قيمة α ٠,١٠ ، وقيم α الأكثر استخداما هي : ٠,٠١ ، ٠,٠٢٥ ، ٠,٠٥ ، ٠,١٠ ،

وتسمى حالة عدم رفض فرض العدم بينما هو في الواقع غير صحيح بالخطأ من النوع الثاني type II error أو خطأ β (الحرف الإغريقي β) β error . ففي مثال المصابيح الكهربائية يكون متوسط عمر المصباح في المصنع أقل من ١٠٠٠ ساعة ، ولكن نتيجة للصدفة فإن العينة العشوائية المسحوبة من هذا المجتمع كان متوسط عمر المصباح فيها أكبر من أو يساوي ١٠٠٠ ساعة ، ومن ثم فإننا نتخذ قرار خاطئ بعدم رفض H_0 . وتمثل β احتمال حدوث الخطأ من النوع الثاني ، أي احتمال عدم رفض فرض العدم H_0 بينما هو في الواقع غير صحيح . وتسمى القيمة (١ - β) بقوة الاختبار power of the test وهي تمثل احتمال عدم حدوث الخطأ من النوع الثاني . ويبين جدول (١) جميع الحالات الممكنة عند إجراء أي اختبار إحصائي :

القرار \ الواقع	H_0 صحيح	H_0 غير صحيح
رفض H_0	α Type I error	$(1 - \beta)$ قرار سليم
عدم رفض H_0	$(1 - \alpha)$ قرار سليم	β Type II error

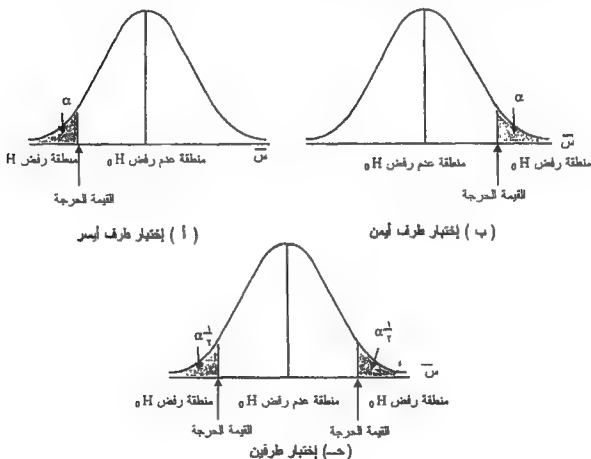
جدول (١) : جميع الحالات الممكنة عند إجراء أي اختبار إحصائي

كلما زادت قيمة α كلما زاد احتمال حدوث الخطأ من النوع الأول أي كلما زاد احتمال رفض فرض العدم بينما هو صحيح . وكلما زادت قيمة β كلما زاد احتمال حدوث الخطأ من النوع الثاني ، أي كلما زاد احتمال عدم رفض فرض العدم بينما هو غير صحيح . ويعتمد كلا من الخطأ من النوع الأول والخطأ من النوع الثاني على الآخر . فبالنسبة لحجم عينة معي ، فلا يمكن تخفيض كلا من α ، β معا في أي اختبار إحصائي : فتخفيض قيمة α يؤدي إلى زيادة قيمة β ، وبالمثل تخفيض قيمة β يؤدي إلى زيادة قيمة α . ولكن هناك طريقة وحيدة يمكن بمقتضاها تخفيض كل من α ، β ألا وهي زيادة حجم العينة .

ويلاحظ في تحليلنا السابق أننا لم نذكر عبارة " قبول فرض العدم " ، لأن كلمة قبول تحمل في طياتها أن هناك قدر كبير من اليقين ، بل ذكرنا بدلا منها عبارة " عدم رفض فرض العدم " .

ولقد سبق وذكرنا عند صياغة الفروض الإحصائية أن هناك ثلاثة أنواع من الاختبارات : اختبار طرف أيسر ، واختبار طرف أيمن ، واختبار طرفين . وعند إجراء أي اختبار من هذه الاختبارات للثلاث نجد أن هناك نقطة على المحور الأفقي للتوزيع اعني تسمى للقيمة الحرجة critical value ، وتحدد

هذه القيمة من جدول التوزيع الاحتمالي لإحصائية الاختبار . وتقوم القيمة الحرجة بتقسيم المحور الأفقي للتوزيع العيني إلى منطقتين : الأولى حول مركز التوزيع تسمى بمنطقة عدم رفض H_0 ، والثانية هي منطقة رفض H_0 ، وهي المنطقة التي تكون مساحتها تحت المنحنى تساوي مستوى المعنوية α . وتسمى منطقة رفض H_0 بالمنطقة الحرجة critical region . ويبين شكل (١) مناطق رفض وعدم رفض H_0 ، بالنسبة لاختبار الطرف الأيسر ، وبالنسبة لاختبار الطرف الأيمن والنسبة للطرفين .



شكل (١) : مناطق رفض وعدم رفض H_0 في حالة :

(أ) اختبار طرف أيسر (ب) اختبار طرف أيمن (جـ) اختبار طرفين

٤ - اتخاذ القرار الإحصائي :

يمكن اتخاذ القرار الإحصائي بإحدى طريقتين رئيسيتين : الأولى باستخدام القيمة الحرجة والثانية باستخدام مستوى الدلالة المشاهد $observed$ level of significance أي القيمة الاحتمالية لإحصائية الاختبار p -value .

أ - باستخدام القيمة الحرجة :

يتم اتخاذ القرار الإحصائي عن طريق مقارنة القيمة المحسوبة لإحصائية الاختبار بالقيمة الحرجة المستخرجة من جدول التوزيع الإحصائي لإحصائية الاختبار . فإذا كانت القيمة العددية لإحصائية الاختبار تزيد عن القيمة الحرجة ، أي إذا وقعت قيمة إحصائية الاختبار في منطقة رفض H_0 ، فإننا نرفض فرض العدم بمستوى المعنوية α الذي تم تحديده مقدماً . أما إذا كانت القيمة العددية لإحصائية الاختبار تقل عن القيمة الحرجة ، أو بمعنى آخر إذا وقعت إحصائية الاختبار في منطقة عدم رفض H_0 ، فإننا لا نرفض فرض العدم بالمستوى المعنوية α المحدد مقدماً . وإذا كانت إحصائية الاختبار تساوي القيمة الحرجة ، فإننا لا نرفض فرض العدم بمستوى المعنوية α .

وفيما يلي بعض القيم الحرجة الأكثر استخداماً في اختبارات الفروض

الإحصائية :

القيمة الحرجة Z	نوع الاختبار	مستوى المعنوية α
$2,326 = \dots Z$	طرف أيمن	$0,01$
$2,326 = \dots Z-$	طرف أيسر	$0,01$
$2,575 = \dots Z$	طرفين	$0,01$
$1,645 = \dots Z$	طرف أيمن	$0,05$
$1,645 = \dots Z-$	طرف أيسر	$0,05$
$1,96 = \dots Z$	طرفين	$0,05$

ب - باستخدام القيمة الاحتمالية لإحصائية الاختبار : p-value

لقد سبق وذكرنا أن مستوى المعنوية α يحدد قبل القيام بالاختبار الإحصائي ، وأنه يجب أن يكون صغيراً لا يتعدى ٠,١ . ألا أن اختيار مستوى المعنوية متروك للقائم بالبحث الإحصائي ، فقد يختار أحد الباحثين مستوى معنوية $\alpha = ٠,٠٥$ ، بينما يفضل باحث آخر اختيار مستوى معنوية $\alpha = ٠,٠١$. ومن ثم ، باستخدام نفس البيانات فإن أحد الباحثين يستنتج أنه يجب رفض فرض العدم بمستوى معنوية ٠,٠٥ ، بينما يستنتج الباحث الآخر بعدم رفض فرض العدم بمستوى معنوية ٠,٠١ . وحتى مستويات المعنوية المستخدمة ٠,٠١ أو ٠,٠٥ فإنها تحدد بطريقة تحكمية أو بالعادة .

ويمكن تعريف القيمة الاحتمالية p-value لإحصائية اختبار معينة ، بأنها أقل قيمة ممكن أن يأخذها مستوى المعنوية α حتى يتم رفض فرض العدم وفقاً للبيانات المشاهدة . فإذا كانت القيمة الاحتمالية لإحصائية الاختبار كبيرة فإننا لا نرفض فرض العدم ، أما إذا كانت القيمة الاحتمالية لإحصائية الاختبار صغيرة فإننا نرفض فرض العدم .

بالإضافة إلى ذلك يمكن مقارنة القيمة الاحتمالية لإحصائية الاختبار بمستوى المعنوية α المحدد مسبقاً لإجراء الاختبار . فإذا كانت القيمة الاحتمالية لإحصائية الاختبار أكبر من أو تساوي مستوى المعنوية α فإننا لا نرفض فرض العدم H_0 ، وبالعكس إذا كانت القيمة الاحتمالية أقل من مستوى المعنوية α ، فإننا نرفض فرض العدم H_0 .

وخلاصة القول :

لا نرفض فرض العدم H_0 إذا كانت :

قيمة $p \leq$ مستوى المعنوية α

ونرفض فرض العدم H_0 إذا كانت :

قيمة $p >$ مستوى المعنوية α

ويضع كثير من الباحثين في مقالاتهم المنشورة القيمة الاحتمالية لإحصائية الاختبار عند إجراء الاختبارات الإحصائية ، كما نجد أيضاً في مخرجات البرامج الجاهزة ، وهذا يعطي القارئ معلومات أكبر من مجرد قبول أو رفض H_0 بالنسبة لمستوى معنوية معين . وبذلك يستطيع القارئ معرفة إلى أي مدى لا تتفق البيانات المشاهدة مع فرض العدم ، بل أكثر من ذلك فإن كل قارئ يستطيع أن يختار قيمة α التي تؤدي إلى رفض فرض العدم . وهذا الأمر لا يتعارض مع اتخاذ القرار الإحصائي باستخدام القيمة الحرجة .⁽¹⁾

(٣ - ٢) اختبارات الفروض الإحصائية المتعلقة بمتوسط مجتمع في حالة العينات الكبيرة :

رأينا في دراستنا لتقدير معالم المجتمع أنه في حالة العينات الكبيرة ($n \leq 30$) ، وفقاً لنظرية النهاية المركزية ، فإن للتوزيع العيني للوسط الحسابي \bar{x} يكون معتدلاً تقريباً ، سواء كان تباين المجتمع معلوماً أم لا . لذلك فعندما يكون حجم العينة كبيراً ، فإن للمنحنى المعتدل يستخدم في اختبارات الفروض الإحصائية ، ومن ثم فإن إحصائية الاختبار تكون :

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

إذا كانت σ معلومة

(1) Mendenhall W. , Wackerly D. & Scheaffer R. : "Mathematical Statistics with applications." P. WS-Kent Publishing Co. . Boston, Fourth ed., 1990, pp. 447-450.
Neter J. . Wasserman. W. & Whitmore. "Applied statistics". . Allyn & Bason. Boston. 1993. Pp. 230-234.

إذا كانت σ غير معلومة

$$Z = \frac{\bar{m} - \mu}{\frac{s}{\sqrt{n}}}$$

وخلاصة القول :

في حالة العينات الكبيرة :

$$\frac{\bar{m} - \mu}{\frac{s}{\sqrt{n}}} = Z \quad \text{إذا كانت } \sigma \text{ معلومة : إحصائية الاختبار :}$$

(٣)

$$\frac{\bar{m} - \mu}{\frac{s}{\sqrt{n}}} = Z \quad \text{إذا كانت } \sigma \text{ غير معلومة : إحصائية الاختبار :}$$

مثال (١) :

لنأخذ مثال المصابيح الكهربائية الذي تكلمنا عنه في مستهل هذا الفصل .
ففي أحد مصانع إنتاج المصابيح الكهربائية ، أعلنت إدارة الإنتاج أن متوسط
عمر المصباح من إنتاج هذا المصنع هو ١٠٠٠ ساعة بانحراف معياري
٩٠ ساعة . ولقد أراد أحد المستوردين شراء شحنة من إنتاج هذا المصنع فأخذ
عينة عشوائية من ١٠٠ مصباح وقام بإنارتها جميعا فوجد أن متوسط عمر
المصباح $\bar{m} = ٩٧٠$ ساعة .

والمطلوب :

أولا : هل يجب على هذا المستورد قبول أو رفض الشحنة ؟ استخدم مستوى
معنوية ٠,٠١ .

ثانيا : باستخدام القيمة الاحتمالية p ، ما هو قرار المستورد ؟

الحل :

$$\mu = 1000 \text{ ساعة} , \sigma = 90 \text{ ساعة} .$$

$$\bar{x} = 970 \text{ ساعة} , n = 100 \text{ مصباح} , \alpha = 0.01$$

أولاً :

ويتم الحل على ٤ خطوات :

١ - صياغة الفروض الإحصائية :

كما سبق وبيننا فإن فرض العدم يفترض أن القيمة النظرية لمعلمة المجتمع صحيحة ، أي أن $(\mu = 1000)$ ولكن بما أنه من المرغوب فيه أن يكون عمر المصباح أكبر من ١٠٠٠ ساعة ، فإن فرض العدم يكون :

$$H_0 : \mu \leq 1000 \text{ ساعة} .$$

ويكون الفرض البديل :

$$H_1 : \mu > 1000 \text{ ساعة} .$$

والاختبار هنا هو اختبار طرف أيسر

٢ - تعيين إحصائية الاختبار وحسابها :

بما أن $n = 100$ فإن حجم العينة كبير $(n > 30)$ ، ومن ثم باستخدام نظرية النهاية المركزية ، فإن التوزيع العيني للأوساط الحسابية يكون معتدلاً تقريباً بصرف النظر عن شكل التوزيع الأصلي للمجتمع . وتكون إحصائية الاختبار هي :

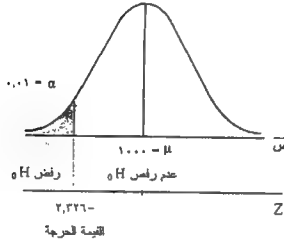
$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$Z = \frac{970 - 1000}{\frac{90}{\sqrt{100}}} = -3.33$$

٣ - تحديد القيمة الحرجة والمنطقة الحرجة :

$$2,326 = Z_{0,01} \quad , \quad 0,01 = \alpha$$

وبما أن الاختبار هو اختبار طرف أيسر فإن القيمة الحرجة هي $Z = -2,326$ وبيّن شكل (٢) القيمة الحرجة والمنطقة الحرجة أي منطقة رفض H_0 .



شكل (٢)

٤ - القرار الإحصائي :

بما أن قيمة Z المصوبة أقل من القيمة الحرجة :

$$-2,326 > -3,33$$

أي أنها تقع في منطقة رفض H_0 ، فإننا نرفض فرض العدم بمستوى معنوية ٠,٠١ ، وهذا يعني أننا نرفض الفرض القائل بأن متوسط عمر المصباح ≤ 1000 ساعة ، أو بمعنى آخر فإن للفرق بين متوسط العينة ومتوسط المجتمع فرق معنوي ، بمستوى معنوية ٠,٠١ . ومن ثم يجب رفض الشحنة .

ثانياً : لقد وجدنا أن قيمة Z المحسوبة = -3.33

نوجد القيمة الاحتمالية p .

بما أن الاختبار طرف أيسر ،

$$\therefore p = P(Z \leq -3.33) = 1 - \Phi(3.33)$$

$$= 1 - 0.9995658$$

$$= 0.0004342$$

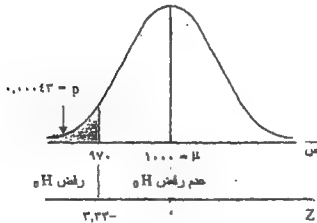
وهنا قيمة p صغيرة جداً $= 0.0004342$ ، أي أنه يتم رفض فرض

العدم بالنسبة لأي مستوى معنوية أكبر من 0.0004342

وبما أن قيمة $p > \alpha$ المستخدمة في الجزء الأول من هذا المثال .

، $0.01 > 0.00043$ ، إذن نرفض H_0 كما سبق وبيننا .

ويبين شكل (٣) قيمة p في هذا المثال .



شكل (٣)

مثال (٢) :

في المثال السابق إذا كان تباين المجتمع غير معلوم ، وكان الانحراف المعياري للعينة ع = ١٠٠ ساعة ، فالمطلوب :

١ — هل يجب على هذا المستورد قبول أو رفض هذه الشحنة ؟

٢ — باستخدام القيمة الاحتمالية ، استنتج قرار المستورد ؟

الحل :

$$١ - \mu = ١٠٠٠ \text{ ساعة}$$

$$\bar{س} = ٩٧٠ \text{ ساعة} ، ع = ١٠٠ \text{ ساعة} ، ن = ١٠٠ \text{ مصباح}$$

الفروض الإحصائية :

$$H_0: \mu \leq ١٠٠٠$$

$$H_1: \mu > ١٠٠٠$$

الاختبار هنا هو اختبار طرف أيسر .

إحصائية الاختبار :

بما أن العينة كبيرة ، فإن التوزيع العيني للأوساط الحسابية يتبع التوزيع المعتدل تقريباً طبقاً لنظرية "نهاية المركزية" ، وبما أن σ غير معلومة نستخدم الانحراف المعياري ع للعينة كمقدر نقطة لمعلمة المجتمع σ . وتكون إحصائية الاختبار هي :

$$Z = \frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}}$$

$$Z = \frac{1000 - 970}{\frac{100}{100\sqrt{}}} = 3$$

القيمة الحرجة :

$$2,326 = Z_{0,01} \quad , \quad \alpha = 0,01$$

القيمة الحرجة : $Z = 2,326$ لأنه اختبار طرف أيسر ، والقيمة الحرجة والمنطقة الحرجة هنا هي نفسها الموجودة في شكل (٢) .
القرار الإحصائي :

بما أن قيمة Z المحسوبة أقل من للقيمة الحرجة :

$$2,326 < 3$$

أي أنها تقع في منطقة رفض H_0 ، فإننا نرفض فرض العدم بمستوى معنوية ٠,٠١ . وهذا يعني أن الفرق بين متوسط للعينة ومتوسط المجتمع فرق معنوي (أي حقيقي) ولا يمكن إرجاعه للصدفة . ومن ثم فإنه يجب رفض الشحنة .

$$2 - \text{ لقد وجدنا أن قيمة } Z \text{ المحسوبة } = 3$$

$$\therefore p = P(Z \geq 3) = 1 - \Phi(3)$$

$$= 1 - 0,99875 = 0,00125$$

قيمة p صغيرة جدا ، أي أن أقل قيمة ممكنة يمكن أن يأخذها مستوى المعنوية حتى يتم رفض فرض العدم هي ٠,٠٠١٢٥ .

وبما أن : $0,00125 < 0,01$ ، إذن نرفض H_0 كما سبق واستنتجنا في الجزء الأول من الإجابة .

مثال (٣) :

طبقا لإحصاءات التعداد في أحد المجتمعات وجد أن متوسط دخل الأسرة لا يتعدى ٢٠٠٠٠ دولار ، ولقد قام فريق من الباحثين بأخذ عينة عشوائية من ٥٠ أسرة فوجد أن متوسط دخل الأسرة هو ٢٠٥٠٠ دولار بانحراف معياري ٢٠٠٠ دولار . وبناءا على ذلك استنتج هذا الفريق بأن بيانات التعداد الخاصة بالدخل ليست دقيقة وأن دخل الأسرة أكبر مما جاء في التعداد . هل تتفق مع هذا الفريق في الرأي ؟ استخدم مستوى معنوية ٠.٠١ .

الحل :

$$\mu = 20000$$

$$\bar{s} = 20500 , \quad c = 2000 , \quad n = 50$$

الفروض الإحصائية :

$$H_0: \mu \geq 20000$$

$$H_1: \mu < 20000$$

الاختبار هو اختبار طرف أيمن . لأن رأس التمرين هنا يعطي فرض H_0 ، حيث نذكر أن متوسط دخل الأسرة لا يتعدى ٢٠٠٠ دولار (أي $\mu \geq 20000$) ، ومن ثم فإن للفرض البديل يكون : $H_1: \mu < 20000$ ويكون الاختبار هو اختبار طرف أيمن .

إحصائية الاختيار :

بما أن العينة كبيرة $n = 50 > 30$ ، فوفقا لنظرية النهاية المركزية فإن توزيع المعاينة للوسط الحسابي \bar{x} يكون معتدلا تقريبا ، ونظرا لأن σ غير

معلومة نستخدم الانحراف المعياري للعينة كمقدر نقطة لها . وتصبح إحصائية الاختبار :

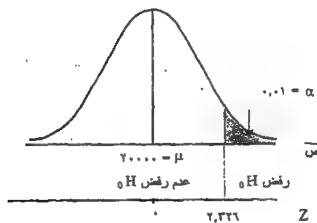
$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = Z$$

$$1,768 = \frac{20000 - 20500}{\frac{2000}{50\sqrt{}}} = Z$$

تحديد القيمة الحرجة والمنطقة الحرجة :

$$\alpha = 0,01 = Z_{0,01} = 2,326 = \text{القيمة الحرجة}$$

ويبين شكل (٤) القيمة الحرجة والمنطقة الحرجة .



شكل (٤)

القرار الإحصائي :

بما أن قيمة Z المصوبة > للقيمة الحرجة :

$$2,326 > 1,768$$

أي أنها تقع في منطقة عدم رفض H_0 ، ومن ثم فإننا لا نرفض فرض عدم H_0 بمستوى معنوية ٠,٠١ ، أي أنه ليس هناك فرق معنوي بين متوسط المجتمع ومتوسط العينة . ومن ثم فإننا لا ننفي في الرأي مع الفريق بأن دخل الأسرة أكبر مما جاء في التعداد .

مثال (٤) :

ينتج أحد المصانع نوع معين من المسامير قطره ٢ ملليمتر . وأراد أحد التجار شراء شحنة من هذه المسامير فأخذ عينة عشوائية من ٤٠ مسمار فوجد أن الوسط الحسابي لقطر المسمار ١,٨ ملليمتر بانحراف معياري ٠,٣ ملليمتر . ولقد استنتج التاجر بأن هذه المسامير غير مطابقة للمواصفات . فهل تتفق معه في الرأي ؟ استخدم مستوى معنوية ٠,٠٥ .

الحل :

$$\mu = 2 \text{ ملليمتر}$$

$$\bar{x} = 1,8 \text{ ملليمتر} , s = 0,3 \text{ ملليمتر} , n = 40$$

الفروض الإحصائية :

$$H_0 : \mu = 2$$

$$H_1 : \mu \neq 2$$

الاختبار هنا اختبار طرفين لأنه بالنسبة للمسامير فإنه من غير المستحب أن يزيد أو يقل قطر المسامير عن ٢ ملليمتر .

إحصائية الاختبار :

العينة هنا $= 40 < 30$ ، أي عينة كبيرة ، وطبقاً لنظرية النهاية المركزية فإن التوزيع العيني للأوساط الحسابية يكون معتدلاً تقريباً ، ولما كانت σ غير معلومة فإننا نستخدم الانحراف المعياري للعينة ع كمقدر نقطة لها .
وتصبح إحصائية الاختبار :

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = Z$$

$$-4,216 = \frac{2 - 1,8}{\frac{0,3}{\sqrt{40}}} = Z$$

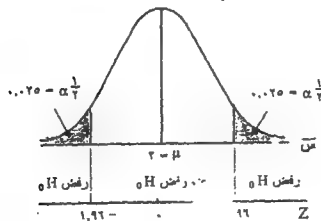
تحديد القيمة الحرجة والمنطقة الحرجة :

$$1,96 = 0,05Z \quad , \quad 0,05 = \alpha \frac{1}{2} \quad , \quad 0,10 = \alpha$$

بما أن الاختبار هو اختبار طرفين ، لذلك فإن هناك قيمتين حرجتين هما :

$$1,96+ \quad , \quad 1,96-$$

ويبين شكل (٥) القيمتين الحرجتين والمنطقة الحرجة .



شكل (٥)

القرار الإحصائي :

بما أن قيمة $|Z|$ المحسوبة < القيمة الحرجة :

$$1,96 > |4,216|$$

أي أن قيمة Z تقع في منطقة رفض H_0 ، فإننا نرفض فرض العدم بمستوى معنوية ٠,٠٥ . ونتفق مع التاجر في الرأي بأن المسامير غير مطابقة للمواصفات .

العلاقة بين اختبارات الفروض وفترات الثقة :

بالإضافة إلى ما سبق يمكننا إجراء اختبارات الفروض عن طريق فترات الثقة — التي قمنا بدراستها في الفصل السابق — ولإيضاح ذلك نأخذ مثال (٤) .

مثال (٤) : وهو اختبار طرفين :

$$\mu = \mu_0 : H_0$$

$$\mu \neq \mu_0 : H_1$$

وبدلاً من اتباع طريقة الحل التي سبق وقمنا بها في حل هذا المثال ، فيمكننا اختبار هذه الفروض بتكوين فترة ثقة $(1 - \alpha) \%$ للمعلمة μ . فإذا كانت هذه الفترة تحتوي على القيمة $\mu = \mu_0$ فإننا لا نرفض H_0 . وبالعكس إذا كانت هذه الفترة لا تحتوي على القيمة $\mu = \mu_0$ فإننا نرفض H_0 .

وفي مثالنا هذا بما أن n كبيرة وطبقاً لنظرية النهاية المركزية فإن التوزيع العيني للأوساط الحسابية من يتبع للتوزيع المعتدل . وهنا مستوى المعنوية $\alpha = 0,05$ ، إذن تقدير فترة ثقة ٩٥ % للوسط الحسابي للمعلمة μ هو :

$$\frac{\bar{c}}{\sqrt{n}} \approx Z \pm \bar{m}$$

$$\left(\frac{0.3}{4.0\sqrt{}} \right) 1.96 \pm 1.8$$

$$0.093 \pm 1.8$$

الحد الأدنى لفترة الثقة = 1.707

الحد الأعلى لفترة الثقة = 1.893

وبما أن فترة الثقة لا تحتوي على القيمة $\mu = 2$ ، إذن نرفض فرض العدم بمستوى معنوية 0.05 وهو نفس القرار الإحصائي الذي توصلنا إليه عند الحل بطريقة القيمة الحرجة .

حساب الخطأ من النوع الثاني β :

عرفنا - في مستهل هذا الفصل - الخطأ من النوع الثاني بأنه احتمال عدم رفض فرض العدم H_0 بينما هو في الواقع غير صحيح . ولحساب الخطأ من النوع الثاني يجب توافر شرطين : الأول أن فرض العدم غير صحيح ، والثاني أن معلمة المجتمع الحقيقية معلومة . وعادة لا نعرف على وجه اليقين إذا كان فرض العدم صحيحاً أم لا ، وحتى إذا علمنا أن فرض العدم خاطئ فإننا لا نستطيع معرفة القيمة الحقيقية لمعلمة المجتمع . ومن ثم فإن حساب الخطأ من النوع الثاني لا يكون ممكناً في الواقع العملي ⁽¹⁾ . ولإيضاح كيفية حساب الخطأ من النوع الثاني في حالة اختبار طرفين نأخذ المثال التالي .

⁽¹⁾ Mann P. S., op. cit. Pp. 474-475.

مثال (٥) :

نفرض أن فرض العدم في مثال (٤) فرض خاطئ ، وأن الوسط الحسابي الحقيقي لقطر المسامير الذي ينتجها هذا المصنع وقت سحب العينة العشوائية هو $\mu = ١,٩٥$ ملليمتر . فإذا كان مستوى المعنوية $٠,٠٥$ ، أوجد احتمال الحصول على الخطأ من النوع الثاني ثم لحسب قوة الاختبار .

الحل :

$$ع = ٠,٣ ، ن = ٤٠ ، \alpha = ٠,٠٥ ، \frac{1}{\sqrt{n}} = \alpha = ٠,٢٥$$

$$Z_{٠,٢٥} = ١,٩٦$$

$$\mu : {}_0H$$

$$\mu \neq : {}_1H$$

تحدد المنطقة الحرجة عندما $|Z| < Z_{\frac{\alpha}{2}}$

$$(١) \text{ عندما : } \alpha \frac{1}{\sqrt{n}} Z < \frac{\frac{\mu - \bar{x}}{\frac{\sigma}{\sqrt{n}}}}{\frac{\sigma}{\sqrt{n}}}$$

$$\frac{\sigma}{\sqrt{n}} \alpha \frac{1}{\sqrt{n}} Z + \mu < \bar{x} \quad \text{أي أن :}$$

$$\frac{٠,٣}{\frac{٤,٠}{\sqrt{40}}} ١,٩٦ + ٢ < \bar{x}$$

$$٠,٠٩٣ + ٢ < \bar{x}$$

$$٢,٠٩٣ < \bar{x}$$

$$(٢) \text{ وعندما : } \alpha \frac{1}{4} Z - > \frac{\bar{M} - \bar{m}}{\frac{\bar{C}}{\sqrt{N}}}$$

$$\frac{\bar{C}}{\sqrt{N}} \alpha \frac{1}{4} Z - \mu > \bar{M}$$

$$\frac{٠,٣}{\sqrt{٤٠}} ١,٩٦ - ٢ > \bar{M}$$

$$٠,٠٩٣ - ٢ > \bar{M}$$

$$١,٩٠٧ > \bar{M}$$

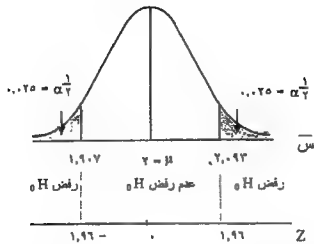
ومن ثم فإن قيمتي \bar{M} والمقابلتين للقيمتين الحرجتين -١,٩٦ ، ١,٩٦ هما ١,٩٠٧ ، ٢,٠٩٣ على التوالي كما هو موضح في شكل (٦ - أ) . ثم نقوم برسم التوزيع العيني للأوساط الحسابية \bar{M} بأخذ الوسط الحسابي الحقيقي للمجتمع ($\mu = ١,٩٥$) ثم نقوم بحساب الخطأ من النوع الثاني β .

وبين شكل (٦ - أ) التوزيع العيني للأوساط الحسابية عندما كانت معلمة المجتمع $\mu = ٢$. كما بين شكل (٦ - ب) التوزيع العيني للأوساط الحسابية عندما كانت معلمة المجتمع $\mu = ١,٩٥$. وتبين المساحة المظلمة تحت المنحنى في الشكل (٦ - ب) للخطأ من النوع الثاني . ومن الملاحظ أن هذه المساحة تقابل منطقة عدم رفض H_0 على المحور الأفقي في شكل (٦ - أ) . ويمكن حساب هذه المساحة - أي الخطأ من النوع الثاني β - باتباع الخطوات التالية :

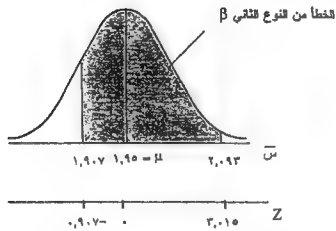
١ - إيجاد قيمتي Z في شكل (٦ - ب) التي تتأخر قيمتي \bar{M} : ١,٩٠٧ ، ٢,٠٩٣

٢ - إيجاد المساحة بين هاتين القيمتين على المنحنى المعتدل المعياري . وقيمة هذه المساحة هي احتمال الحصول على الخطأ من النوع الثاني β .

(٦ - أ)



(٦ - ب)



شكل (٦)

١ - عند $\bar{x} = 1.907$

$$Z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \text{القيمة المعيارية}$$

$$-1.907 = \frac{1.90 - 1.907}{\frac{0.3}{\sqrt{40}}}$$

عند $\bar{s} = 2,093$

$$3,010 = \frac{1,90 - 2,093}{\frac{0,3}{\sqrt{40}}} = Z \quad \text{القيمة المعيارية :}$$

$$\beta = P(3,010 > Z > 0,907)$$

$$= \Phi(0,907) - \Phi(3,010)$$

$$= [\Phi(0,907) - 1] - [\Phi(3,010) - 1]$$

$$= (0,8186 - 1) - (0,0044) =$$

$$= -0,1858$$

وهذا يعني أن احتمال عدم رفض فرض العدم بينما هو في الواقع غير

$$\text{صحيح} = 0,1858$$

$$\text{قوة الاختبار} = 1 - \beta$$

$$= 1 - 0,1858$$

$$= 0,8142$$

ومن ثم فإن احتمال رفض فرض العدم بينما هو في الواقع غير صحيح

$$\text{هو } 0,1858$$

مثال (٦) :

ولنفرض في مثال (٣) أن فرض العدم غير صحيح وأن متوسط دخل الأسرة في المجتمع هو ٢٠٣٠٠ دولار ، فإذا كان مستوى المعنوية ٠,٠١ أوجد احتمال الحصول على الخطأ من النوع الثاني ثم احسب قوة الاختبار .

الحل :

$$ع = ٢٠٠٠٠ ، ن = ٥٠ ، \alpha = ٠,٠١ ، Z_{\alpha} = ٢,٣٢٦$$

$$H_0: \mu \geq ٢٠٠٠٠$$

$$H_1: \mu < ٢٠٠٠٠$$

الاختبار هنا هو اختبار طرف أيمن .

لذلك فإن المنطقة الحرجة تكون عندما $Z < Z_{\alpha}$

$$\text{أي : } Z < \frac{\bar{X} - \mu}{\frac{ع}{\sqrt{ن}}}$$

$$\bar{X} < \frac{ع}{\sqrt{ن}} Z_{\alpha} + \mu$$

$$\bar{X} < ٢,٣٢٦ + \frac{٢٠٠٠}{٥٠\sqrt{}}$$

$$\bar{X} < ٢٠٦٥٧,٨٩٢$$

إن قيمة \bar{S} المقابلة للقيمة الحرجة $Z = 1.1$ هي $\bar{S} = 2.326$ ، $2.657, 892 = \bar{S}$ ويبين شكل (٧ - أ) التوزيع العيني للأوساط الحسابية عندما كانت معلومة المجتمع $\mu = 2.000$ ، مبينا للقيمة الحرجة والقيمة المقابلة لها من قيم \bar{S} .

نوجد قيمة Z التي تتأظر قيمة $\bar{S} = 2.657, 892$ بالنسبة للتوزيع العيني الجديد ، حيث $\mu = 2.300$ كما هو مبين في شكل (٧ - ب) .

$$\frac{\mu - \bar{S}}{\frac{\sigma}{\sqrt{n}}} = Z$$

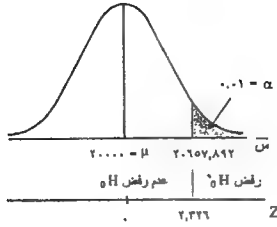
$$\frac{2.300 - 2.657, 892}{\frac{2.000}{50\sqrt{}}} = 1.265 =$$

ويمكن حساب الخطأ من النوع الثاني β بأنه المساحة على يسار قيمة $Z = 1.265$ تحت منحنى التوزيع العيني الجديد في شكل (٧ - ب) .
أي أن :

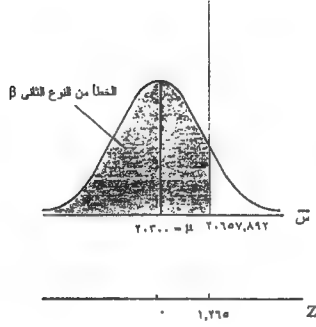
$$\beta = 1 - \Phi(1.265) = 0.8971 =$$

أي أن احتمال عدم رفض فرض العدم بينما هو غير صحيح = 0.8971 .

(٧ - ١)



(٧ - ب)



شكل (٧)

وتحزن قوة الاختبار $1 - \beta$

$$0.8971 = 1 - 0.1029$$

أي أن احتمال رفض H_0 بينما هو غير صحيح 0.1029

(٣ - ٣) اختبارات الفروض الإحصائية المتعلقة بمتوسط

مجتمع في حالة العينات الصغيرة :

كما رأينا سلفاً في الفصل السابق - في مبحث (٢ - ٤) عند دراسة تقدير فترة ثقة لمتوسط مجتمع في حالة العينات الصغيرة - إذا كانت العينات صغيرة ($n > 30$) يجب التفرقة بين حالة ما إذا كانت σ معلومة أم غير معلومة .

أ - إذا كانت σ معلومة : وكان التوزيع الأصلي للمجتمع الذي أخذت منه العينة هو توزيعاً معتدلاً أو قريباً من الاعتدال ، فإن التوزيع العيني للأوساط الحسابية يكون أيضاً معتدلاً ، ومن ثم فإن إحصائية الاختيار في هذه الحالة هي :

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = Z$$

ب - إذا كانت σ غير معلومة : وكان التوزيع الذي سحبت منه العينة توزيعاً معتدلاً أو قريباً من الاعتدال ، فإننا نستخدم الانحراف المعياري للعينة \bar{s} كمقدر نقطة للمعلمة σ ، فإن إحصائية الاختيار في هذه الحالة هي :

$$\frac{\bar{x} - \mu}{\frac{\bar{s}}{\sqrt{n}}} = t$$

وهي تتبع توزيع t .

وخلاصة القول :

$$(٦) \quad \begin{aligned} &\text{إذا كانت } \sigma \text{ معلومة : إحصائية الاختبار : } Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \\ &\text{إذا كانت } \sigma \text{ غير معلومة : إحصائية الاختبار : } t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \end{aligned}$$

ولإجراء اختبارات الفروض نتبع نفس الطريقة التي اتبعناها عند إجراء اختبارات الفروض في حالة العينات الكبيرة مع اختلاف واحد ألا وهو استخدام إحصائية الاختبار t بدلاً من Z .

مثال (٧) :

ينتج أحد مصانع لعب الأطفال لعبة جديدة تربية عبارة عن مكعبات يقوم الأطفال بتجميعها بأسرع وقت ممكن . ولقد أعلن هذا المصنع أن متوسط الوقت الذي يقضيه طفل عمره ٥ سنوات لتجميع هذه المكعبات هو ٢٥ دقيقة على الأقل . ويتبع توزيع وقت تجميع المكعبات التوزيع المعتدل . ولقد ادعت إحدى المدارس الخاصة بأن تلاميذها الذين يبلغون من العمر ٥ سنوات متميزون ويتمتعون بقدرات خاصة وأن الوقت الذي يقضونه في تجميع هذه المكعبات أقل من ٢٥ دقيقة . ولقد أخذت هذه المدرسة عينة عشوائية من ٢٠ تلميذ من فئة عمر ٥ سنوات ، وكان متوسط الوقت المنقضي في تجميع هذه المكعبات هو ٢٢ دقيقة بانحراف معياري ٢,٥ دقيقة . فهل تتفق مع رأي هذه المدرسة ؟ استخدم مستوى معنوية $\alpha = ٠,٠١$.

الحل :

$$\mu = 20 \text{ دقيقة} , \quad \bar{x} = 22 \text{ دقيقة} , \quad \sigma = 2,0 \text{ دقيقة}$$

$$n = 20 , \quad \alpha = 0,01$$

الفروض الإحصائية :

$$H_0 : \mu \leq 20$$

$$H_1 : \mu > 20$$

وهو اختبار طرف أيسر ، لأن رأس التمرين ينص على أن متوسط الوقت الذي يقضيه الطفل لتجميع المكعبات هو 20 دقيقة على الأقل أي أن $H_0 : \mu \leq 20$ ، ويكون $H_1 : \mu > 20$.

إحصائية الاختبار :

العينة هنا صغيرة ($n > 30$) ، والمجتمع الذي سحبت منه العينة يتوزع توزيعاً معتدلاً ، والانحراف المعياري للمجتمع غير معلوم ، لذلك فإن التوزيع العيني للأوساط الحسابية \bar{x} يخضع لتوزيع t ، وتكون إحصائية الاختبار هي :

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{22 - 20}{\frac{2,0}{\sqrt{20}}} = 0,3666$$

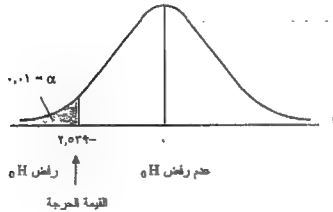
تحديد القيمة الحرجة والمنطقة الحرجة :

$$\alpha = 0.01 \quad , \quad n - 1 = 19$$

$$\text{القيمة الحرجة} = t_{(0.01, 19)} = 2.539$$

وتكون المنطقة الحرجة حيث : $t > t_{(0.01, 19)}$

كما هو مبين في شكل (٨)



شكل (٨)

القرار الإحصائي :

بما أن قيمة t المحسوبة أقل من القيمة الحرجة :

$$2.539- > 0.3666-$$

أي أنها تقع في منطقة رفض H_0 ، فإننا نرفض فرض العدم بمستوى معنوية ٠,٠١ . وهذا يعني أننا نتفق مع ادعاء المدرسة بأن تلاميذها تأخذ وقت أقل في جميع المكعبات ، وهذا بمستوى معنوية ٠,٠١ .

مثال (٨) :

أعلن أحد مديري شركات الطيران أن متوسط عدد الأماكن الشاغرة على أحد خطوط رحلات الطيران لا يتعدى ١٠ مقاعد . ولمعرفة مدى صحة هذا الإدعاء قامت إدارة البحوث باختيار عينة عشوائية من ١٦ رحلة طيران على هذا الخط فوجد أن الوسط الحسابي ١٢,٧ مقعد والانحراف المعياري ٤,٥ مقعد . فإذا علمت أن توزيع الأماكن الشاغرة في الطائرات يتبع التوزيع المعتمد ، فهل تتفق مع مدير شركة الطيران في الرأي ؟ استخدم مستوى معنوية ٠,٠٥ .

الحل :

$$\mu = 10 , \quad n = 16 , \quad \bar{x} = 12.7 , \quad s = 4.5 , \quad \alpha = 0.05$$

الفروض الإحصائية :

$$H_0 : \mu \geq 10$$

$$H_1 : \mu < 10$$

هذا الاختبار هو اختبار طرف أيمن ، لأن رأس التمرين ينحصر على أن عدد الأماكن الشاغرة لا يتعدى ١٠ مقاعد أي أن $H_0 : \mu \geq 10$ ، ويكون $H_1 : \mu < 10$.

إحصائية الاختبار :

العينة هنا صغيرة ($n > 30$) ، وللتوزيع الذي محبت منه العينة توزيعاً معتملاً ، والانحراف المعياري للمجتمع غير معلوم ، لذلك فإن التوزيع العيني للأوساط الحسابية \bar{x} يخضع لتوزيع t ، وتكون إحصائية الاختبار هي :

$$\frac{\mu - \bar{m}}{\frac{\bar{s}}{\sqrt{n}}} = t$$

$$2,4 = \frac{10 - 12,7}{\frac{4,5}{\sqrt{16}}}$$

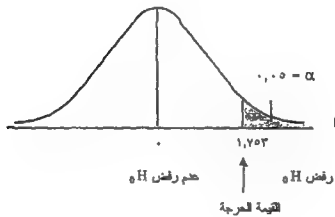
تحديد القيمة الحرجة والمنطقة الحرجة :

$$10 = 1 - \alpha \quad , \quad 0,05 = \alpha$$

القيمة الحرجة $t = (0,05, 10) = 1,753$

وتكون المنطقة الحرجة حيث : $t < (0,05, 10)$

كما هو مبين في شكل (٩)



شكل (٩)

القرار الإحصائي :

بما أن قيمة t المحسوبة أكبر من القيمة الحرجة :

$$1,753 < 2,4$$

أي أنها تقع في منطقة رفض H_0 ، فإننا نرفض H_0 بمستوى معنوية ٠,٠٥ . وهذا يعني أننا نرفض ادعاء المدير بأن متوسط عدد الأماكن الشاغرة على هذا الخط من الطيران لا يتعدى ١٠ مقاعد .
مثال (٩) :

في مشروع التخرج قام خمسة من الطلاب بقياس مساحة قطعة أرض . وكانت المساحة حسب مقاس كل منهم بالفدان هي :

٥,٢٦ ، ٥,٢٣ ، ٥,٢١ ، ٥,٢٧ ، ٥,٢٢

إذا علمت أن توزيع قياسات مساحة الأرض هو توزيع معتدل ، وأن مسالك الأرض أعلن أن المساحة الحقيقية لهذه الأرض هي ٥,٢٢ فدان ، المطلوب : اختبار الفرض القائل بأن المساحة الحقيقية لقطعة الأرض هي ٥,٢٢ فدان ، استخدم مستوى معنوية ٠,٠٥ .

الحل :

$$\mu = ٥,٢٢ \text{ فدان} ، \alpha = ٠,٠٥$$

من بيانات العينة نحسب أولاً الوسط الحسابي \bar{x} والانحراف المعياري s .

المساحة (بالفدان) س	س - \bar{x}	(س - \bar{x}) ^٢
٥,٢١	-٠,٠٢٨	٠,٠٠٠٧٨٤
٥,٢٢	-٠,٠١٨	٠,٠٠٠٣٢٤
٥,٢٣	-٠,٠٠٨	٠,٠٠٠٠٦٤
٥,٢٦	٠,٠٢٢	٠,٠٠٠٤٨٤
٥,٢٧	٠,٠٣٢	٠,٠٠١٠٢٤
٢٦,١٩		٠,٠٠٢٦٨٠

$$\bar{x} = \frac{\sum s}{n} = \frac{٢٦,١٩}{٥} = ٥,٢٣٨$$

$$s^2 = \frac{1}{n-1} \sum (s - \bar{x})^2$$

$$0,00067 = (0,00268) \frac{1}{1-0} =$$

$$ع = 0,26 = \sqrt{0,00067} \text{ فدان}$$

الفروض الإحصائية :

$$0,22 = \mu : {}_0H$$

$$0,22 \neq \mu : {}_1H$$

والاختبار هنا اختبار طرفين .

إحصائية الاختبار :

العينة هنا صغيرة ($n = 30 > 0$) ، المجتمع الذي سحبت منه هذه العينة يتبع التوزيع المعتدل ، والانحراف المعياري للمجتمع غير معلوم ، لذلك فإن التوزيع العيني للأوساط الحسابية من يخضع لتوزيع t ، وتكون إحصائية الاختبار هي :

$$t = \frac{\bar{x} - \mu}{\frac{ع}{\sqrt{n}}}$$

$$1,048 = \frac{0,22 - 0,238}{\frac{0,26}{\sqrt{30}}}$$

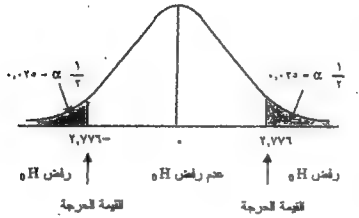
تحديد القيمة الحرجة والمنطقة الحرجة :

$$\alpha = 0,05 \quad , \quad \frac{1}{\alpha} = 0,20 \quad , \quad \alpha = 1 - \alpha$$

بما أن الاختبار هو اختبار طرفين ، إذن هناك قيمتين حرجتين وهما :

$$2,776 \pm (0,050,4) t \pm$$

وتكون المنطقة للدرجة حيث : $|t| < (0,050,4) t$ كما هو مبين في شكل (١٠)



شكل (١٠)

القرار الإحصائي :

بما أن قيمة $|t| >$ لقيمة الدرجة

$$2,776 > |1,048|$$

أي أن t تقع في منطقة عدم رفض H_0 ، إذن لا نرفض فرض العدم بمستوى معنوية ٠,٠٥ أي أننا لا نرفض إعلان مالك الأرض بأن المساحة الحقيقية لقطعة الأرض هي ٥,٢٢ فدان .

(٣ - ٤) اختبارات الفروض الإحصائية المتعلقة بنسبة

مجتمع في حالة العينات الكبيرة :

سبق وتناولنا في المبحث (٢ - ٥) تقدير فترة الثقة للنسبة في المجتمع حيث اقتصرنا دراستنا على العينات الكبيرة حيث $0 \leq \theta$. وطبقا لنظرية النهاية المركزية فمع زيادة حجم العينة فإن للتوزيع العيني يقترب من التوزيع المعتدل الذي متوسطه θ ، وتباينه $\frac{\theta(1-\theta)}{n}$.

وبالنسبة لاختبارات الفروض الإحصائية المتعلقة بنسبة مجتمع في حالة العينات الكبيرة ، فإنها تشبه إلى حد كبير اختبارات الفروض المتعلقة بمتوسط مجتمع في حالة العينات الكبيرة . لذلك سنتبع نفس المراحل الأربعة السابق استخدامها في اختبارات الفروض . فبالنسبة لصياغة الفروض الإحصائية المتعلقة بالنسبة θ ، بافتراض أن θ_0 هي قيمة معلمة المجتمع θ :

فعند اختبار طرف أيسر :

$$H_0 : \theta \leq \theta_0$$

$$H_1 : \theta > \theta_0$$

فعند اختبار طرف أيمن :

$$H_0 : \theta \geq \theta_0$$

$$H_1 : \theta < \theta_0$$

وعند اختبار طرفين :

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

أما بالنسبة لإحصائية الاختبار فستكون :

(٧)

$$\frac{q - \theta_0}{\sqrt{\frac{\theta_0(1 - \theta_0)}{n}}} = Z$$

حيث : θ = قيمة للنسبة في المجتمع .

$q = \hat{\theta}$ = النسبة في العينة

n = حجم العينة

ويتم تحديد القيمة الحرجة من جدول المنحنى المعتدل المعياري طبقاً لمستوى المعنوية α ، فتكون Z_{α} في حالة لختبار الطرف الأيسر ، Z_{α} في حالة اختبار الطرف الأيمن ، $Z_{\frac{\alpha}{2}}$ في حالة اختبار الطرفين . ويتم تحديد المنطقة الحرجة بنفس الطريقة التي اتبعناها من قبل ، وكذلك فإن القرار الإحصائي يتم أيضاً بنفس الطريقة السابق اتباعها . ويتضح ذلك من الأمثلة التالية .

مثال (١٠) :

أعلن أحد المسؤولين أن نسبة الأمية في أحد المجتمعات الريفية لا تتعدى ٤٠ % . ولقد قام فريق من الباحثين بسحب عينة عشوائية من ٢٠٠ فرد من هذا المجتمع فوجد أن عدد الأميين فيها ٩٠ شخص ، فهل تؤيد قول المسؤول بأن نسبة الأمية لا تتعدى ٤٠ % ؟ استخدم مستوى معنوية ٠,٠٥ .

الحل :

$$\theta_0 = ٠,٤٠ ، n = ٢٠٠ ، q = \frac{٩٠}{٢٠٠} = ٠,٤٥$$

الفروض الإحصائية :

$$H_0 : \theta \geq ٠,٤٠$$

$$H_1 : \theta < ٠,٤٠$$

الاختبار هنا اختبار طرف ليمين

إحصائية الاختبار :

بما أن العينة كبيرة : $n = \theta = 200 = (0,40) \cdot 500$

$n = (\theta - 1) \cdot 200 = (0,6) \cdot 200$

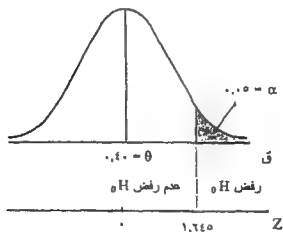
فيمكننا استخدام التوزيع المعتدل ، طبقاً لنظرية النهاية المركزية . وتكون إحصائية الاختبار هي :

$$Z = \frac{q - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} = \frac{0,40 - 0,40}{\sqrt{\frac{(0,40 - 1) \cdot 0,40}{200}}} = 1,443$$

تحديد القيمة الحرجة والمنطقة الحرجة :

$\alpha = 0,05$ ، $Z_{0,05} = 1,645$ = القيمة الحرجة .

ويبين شكل (١١) القيمة الحرجة والمنطقة الحرجة .



شكل (١١)

القرار الإحصائي :

بما أن قيمة Z المحسوبة $>$ للقيمة الحرجة :

$$1,645 > 1,443$$

أي أنها تقع في منطقة عدم رفض H_0 ، فإننا لا نرفض فرض العدم القائل بأن نسبة الأمية لا تتعدى ٤٠ ، بمستوى معنوية $\alpha = ٠,٠٥$. أي أننا نؤيد قول المسئول بأن نسبة الأمية لا تتعدى ٤٠ % .

مثال (١١) :

أرادت إحدى الشركات طرح منتج جديد في الأسواق . ولقد قرر مدير هذه الشركة أنه سيقوم بإنزال هذا المنتج في الأسواق إذا كان هناك على الأقل ٣٣ % من المستهلكين يفضلونه . وتريد الشركة معرفة نسبة المستهلكين الذين يفضلون هذا المنتج ، لذلك قام قسم الأبحاث بالشركة بسحب عينة عشوائية من ٣٠٠ مستهلك وأعطى لهم المنتج الجديد مجاناً . وبعد تجريبه تبين أن ٨٧ منهم يفضلون هذا المنتج . فهل يجب على الشركة تسويق هذا المنتج أم لا ؟ استخدم مستوى معنوية ٠,٠٥ .

الحل :

$$\theta = ٠,٣٣ ، n = ٣٠٠ ، q = \frac{٨٧}{٣٠٠} = ٠,٢٩$$

الفروض الإحصائية :

$$H_0 : \theta \leq ٠,٣٣$$

$$H_1 : \theta > ٠,٣٣$$

والاختبار هنا هو اختبار طرف أيسر .

إحصائية الاختبار :

بما أن العينة كبيرة : $n = 300 = \theta = 0.33$ $0 < 0.1$

$n = 300 = (\theta - 1) = 0.67$ $0 < 0.1$

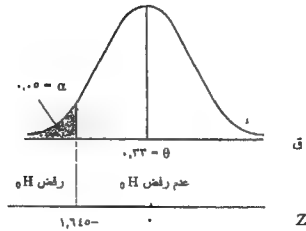
فطبقاً لنظرية النهاية المركزية يمكننا استخدام التوزيع المعتدل . وتكون

إحصائية الاختبار هي :

$$\frac{q - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} = Z$$

$$1.473 = \frac{0.33 - 0.29}{\sqrt{\frac{(0.67) \cdot 0.33}{300}}}$$

تحديد القيمة الحرجة والمنطقة الحرجة :



شكل (١٢)

$$1,645 = Z_{0,05} \quad , \quad 0,05 = \alpha$$

القيمة الحرجة هنا $= 1,645$ لأن الاختبار طرف أيسر ، ويبين شكل (١٢) القيمة الحرجة والمنطقة الحرجة .

القرار الإحصائي :

بما أن قيمة Z المحسوبة أكبر من القيمة الحرجة :

$$1,645 < 1,473$$

أي أنها تقع في منطقة عدم رفض H_0 ، فإننا لا نرفض فرض العدم بمستوى معنوية $\alpha = 0,05$. وهذا يعني أننا لا نرفض الفرض القائل بأن $0,33$ من المستهلكين يفضلون هذا المنتج ، ومن ثم فإن على الشركة القيام بتسويق هذا المنتج .

مثال (١٢) :

أعلنت أحد المحلات أن نسبة المدخنين بين طلبة الجامعة هي 38% . ولقد قام فريق من الباحثين بسحب عينة عشوائية من 400 طالب فوجد أن عدد المدخنين 166 طالب . اختبر الفرض القائل بأن نسبة المدخنين تختلف عن 38% بمستوى معنوية $0,01$.

الحل :

$$\theta = 0,38 \quad , \quad n = 400 \quad , \quad q = \frac{166}{400} = 0,415$$

الفروض الإحصائية :

$$H_0 : \theta = 0,38$$

$$H_1 : \theta \neq 0,38$$

والاختبار هنا هو اختبار طرفين .

إحصائية الاختبار :

$$\text{بما أن العينة كبيرة : } n = \theta = 400 = (0,38) < 100$$

$$n = (\theta - 1) = 399 = (0,62) < 400$$

وطبقاً لنظرية النهاية المركزية يمكننا استخدام التوزيع المعتمد

وتكون إحصائية الاختبار هي :

$$Z = \frac{\bar{q} - q_0}{\sqrt{\frac{q_0(1 - q_0)}{n}}}$$

$$1,442 = \frac{0,38 - 0,415}{\sqrt{\frac{(0,62) \cdot 0,38}{400}}}$$

تحديد القيمة الحرجة والمنطقة الحرجة :

$$2,075 = Z_{\alpha/2} , \quad 0,005 = \alpha/2 , \quad 0,01 = \alpha$$

بما أن الاختبار هو اختبار طرفين ، إذن هناك قيمتين حرجتين هما :

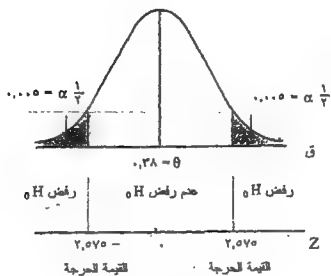
$2,075+$ ، $2,075-$. ويبين شكل (١٣) للقيمتين الحرجتين في الشكل ١٣ .
الدرجة .

القرار الإحصائي :

بما أن قيمة $|Z|$ المحسوبة > القيمة الحرجة :

$$2,075 > |1,442|$$

أي أن قيمة Z المحسوبة تقع في منطقة عدم رفض H_0 ، فإننا لا نرفض فرض
العدم بمستوى معنوية $0,01$. أي أننا لا نرفض الفرض القائل بأن نسبة
المدخنين بين طلبة الجامعة هي $0,38$.



شكل (١٣)

تمارين (٣)

١ - تعاقد أحد مزارع الدولجن على توريد شحنة دجاج لأحد المطاعم .
ويدعي صاحب هذه المزرعة أن متوسط وزن للدجاجة لا يقل عن
١,٢٥ كيلو جرام بانحراف معياري ٠,١٠ كيلو جرام . ولقد قام صاحب
المطعم بسحب عينة عشوائية من ١٠٠ دجاجة فوجد أن متوسط وزنها
١,١٠ كيلو جرام . والمطلوب : .

١ - هل يجب على صاحب المطعم رفض الشحنة ؟ حل باستخدام طريقة
القيمة الحرجة ثم باستخدام طريقة للقيمة الاحتمالية ، استخدم
مستوى معنوية ٠,٠٥ .

٢ - أوجد الخطأ من النوع الثاني واحصب قوة الاختبار إذا علمت أن
فرض العدم غير صحيح وأن متوسط وزن الزجاجة الحقيقي هو
١,٢٤ كيلو جرام .

٢ - في أحد مصانع الإطارات كان العمر الافتراضي لها يتبع التوزيع المعتدل
بمتوسط ٣١٠٠٠ كيلو متر . ولقد قامت إدارة البحوث بالمصنع بإضافة
مادة جديدة لزيادة العمر الافتراضي لهذه الإطارات . وللتأكد من ذلك تم
تصنيع ١٥ إطاراً بإضافة هذه المادة الجديدة ، وبقياس أعمار هذه
الإطارات تبين أن الوسط الحسابي لعمر الإطار هو ٣٣٠٠٠ كيلو متر
بانحراف معياري ٢٠٥٠ كيلو متر . والمطلوب : هل يجب تعميم إضافة
هذه المادة الجديدة على جميع إطارات المصنع ؟ استخدم مستوى
معنوية ٠,٠١ .

٣ - تقوم أحد الآلات بتعينة المياه المعدنية في زجاجات على أن تحتوي
الزجاجة على ١,٥ كيلو جرام من المياه المعدنية ولقد قام مراقب جودة

الإنتاج بسحب عينة من ٢٠ زجاجة فوجد أن الوسط الحسابي والانحراف المعياري لوزن المياه المعدنية في الزجاجات كانا على التوالي : ١,٣٧ كيلو جرام ، ٠,٢٠ كيلو جرام . ولقد استنتج مراقب الجودة أن وزن العبوة مختلف عن المواصفات ، وبافتراض أن المجتمع الذي سحبت منه العينة يتوزع توزيعاً معتدلاً ، فالمطلوب : هل توافق مراقب الجودة في الرأي ؟ استخدم مستوى معنوية ٠,٠٥

٤ - في أحد المصانع تقوم آلة بإنتاج مواسير من الألومنيوم أقطارها ٣ سم . ولقد قام مراقب جودة الإنتاج بسحب عينة عشوائية من ٩ مواسير فوجد أقطارها كالآتي :

٢,٧ ٣,٢ ٣,٤ ٢,٩ ٣,١ ٢,٨ ٣,٣ ٣,٦ ٣,٧

فإذا علمت أن أقطار المواسير يتبع التوزيع الطبيعي تقريباً ، فالمطلوب : باستخدام مستوى معنوية ٠,٠١ تحديد ما إذا كانت هذه الآلة تحتاج إلى صيانة أم لا ؟

٥ - تعلن إحدى الشركات المنتجة لبطاريات السيارات أن العمر الافتراضي للبطارية هو ٤٠ شهر . ولقد قام أحد المستوردين لهذه البطاريات بسحب عينة عشوائية من ٢٠ بطارية فوجد أن متوسط عمر البطارية ٣٧,٨ شهر بانحراف معياري ٣,٥ شهر . وبافتراض أن المجتمع الذي سحبت منه العينة يتوزع توزيعاً معتدلاً ، فالمطلوب : هل توافق على أن عمر هذه البطاريات أقل من ٤٠ شهر ؟ استخدم مستوى معنوية ٠,٠٥

٦ - قامت إحدى شركات المنجائر بإنتاج نوع جديد وأعلنت أن القطر لن يزيد عن ٣,٨ ملليجرام في السجارة . ولقد قام أحد المستوردين بسحب عينة عشوائية من ٢٠ سجارة فوجد أن الوسط الحسابي لكمية القطران بها هو ٤,١٤ ملليجرام بانحراف معياري ع = ٠,٤٠ ملليجرام .

ولقد قام المستورد برفض الشحنة . وبافتراض أن المجتمع الذي سحبت منه هذه العينة يتوزع توزيعاً معتدلاً ، المطلوب : هل تتفق مع المستورد في الرأي ؟ استخدم مستوى معنوية ٠,٠١ .

٧ - تقوم إحدى الشركات بإرسال البريد السريع ، وتعلن هذه الشركة بأنها توصل على الأقل ٦٥ ٪ من البريد في ظرف ٤٨ ساعة . وتقوم إدارة الرقابة على الجودة بالتأكد من وقت إلى آخر بأن البريد يصل في ميعاده . ولقد أخذت هذه الإدارة عينة من ٢٠٠ طرد بريدي ووجدت أن ١٢٠ منها وصلت في خلال ٤٨ ساعة . والمطلوب :

أ - بأخذ مستوى معنوية ٠,٠١ هل تتفق مع الشركة في الرأي ؟
ب - ما هو استنتاجك في (أ) إذا كان احتمال الحصول على الخطأ من النوع الأول يساوي صفر ؟ فسر الإجابة .

٨ - تقوم إحدى شركات الكمبيوتر بإنتاج اسطوانات الكمبيوتر (ديسكات) وتقوم إحدى الآلات بتصنيع هذه الاسطوانات ، ومن المعلوم أن نسبة العادم من إنتاج هذه الآلة لا يتعدى ٥ ٪ . ولقد قام مراقب جودة الإنتاج بسحب عينة عشوائية من ٤٠٠ اسطوانة فوجد أن بها ٣٠ اسطوانة معيبة . والمطلوب :

أ - باستخدام مستوى معنوية ٠,٠٥ ، تحديد ما إذا كانت هذه الآلة تحتاج إلى صيانة .

ب - وباستخدام مستوى معنوية ٠,٠١ ، هل تستنتج نفس الاستنتاج في (أ) ؟

ج - أوجد القيمة الاحتمالية p .

الفصل الرابع

أساليب الاستدلال الإحصائي للمقارنة

بين معالم مجتمعين

Statistical Inference Techniques For Comparing the Parameters Of Two Populations

١-٤ مقدمة

نتناول في هذا الفصل بعض أساليب الاستدلال الإحصائي والتي تناسب ما يسمى بدراسات المقارنة، حيث يكون لدينا في هذه الحالة مجتمعين نهتم بعقد مقارنات بين معالمهما المتناظرة وذلك بهدف التعرف على أوجه الاختلاف والتشابه بينهما. فعلى سبيل المثال، قد نهتم في دراسة معينة بمقارنة أجور العمال وأجور العاملات للتعرف على ما إذا كان هناك تمييز في مستويات الأجور بين النوعين أم لا. كذلك قد نهتم بمقارنة درجات الطلاب والطالبات في امتحان معين، أو مقارنة درجات نفس المجموعة من الطلبة في امتحانين مختلفين. وقد نهتم في دراسة ثالثة بمقارنة تأثير أسلوبين مختلفين لعلاج مرض معين على الحالة الصحية للمصابين بهذا المرض، أو مقارنة الحالة الصحية لمجموعة من المرضى

قبل وبعد تناول دواء معين. ويمكن القول بأن دراسات المقارنة هي الأكثر انتشارا في الحياة العملية نظرا لشعب مجالات تطبيقها.

عند إجراء دراسة تتضمن المقارنة بين مجتمعين، فلا بد في البداية من تحديد العلاقة بين المجتمعين. ونميز في هذا الصدد بين حالتين. الحالة الأولى هي حالة استقلال مشاهدات كل مجتمع عن مشاهدات المجتمع الآخر. والحالة الثانية هي حالة ارتباط كل مشاهدة في المجتمع الأول بمشاهدة مقابلة لها في المجتمع الثاني.

حالة استقلال مجتمعي الدراسة:

سوف نعتبر دائما أن مجتمعي الدراسة مستقلان عن بعضهما البعض إذا كانت مشاهدات أحد المجتمعين لا تتأثر ولا تعتمد على قيم وملاحظات المجتمع الثاني. فعلى سبيل المثال، عند الحديث عن توزيع الدرجات في امتحان معين، يخضع فيه الطلبة لمراقبة جيدة، تعبر درجة كل طالب عن مستواه العلمي ومدى الجهد الذي بذله في التحضير للامتحان، وبالتالي تكون درجات الطلاب مستقلة عن بعضها البعض ومستقلة أيضا عن درجات الطالبات. كذلك عند إعطاء دواء "A" لمرضى بمرض معين، وإعطاء دواء "B" لمرضى آخر بنفس المرض، فإن استجابة المريض الأول للدواء "A" لا تتأثر ولا تعتمد على استجابة المريض الثاني للدواء "B". وبالتالي فإن مجتمع بيانات الحالة الصحية للمرضى الذين تتم معالجتهم بالدواء "A" يكون مستقلا عن مجتمع بيانات الحالة الصحية للمرضى الذين تتم معالجتهم بالدواء "B".

حالة عدم استقلال مجتمعي الدراسة

إذا وجدنا في أية دراسة أن كل مشاهدة في أحد المجتمعين ترتبط بمشاهدة مناظرة في المجتمع الثاني، فإننا نقول بأن مجتمعي الدراسة غير مستقلين ويوجد بينهما ارتباط. وينشأ الارتباط بين مجتمعين للبيانات بأحد أسلوبين أساسيين هما أسلوب القياس على نفس مفردات الدراسة قبل وبعد إخضاع المفردات لمعالجة معينة. وفي الأسلوب الآخر يتم تكوين أزواج متماثلة من المفردات لها تقريبا نفس السمات والخصائص حيث يتم اختيار أحد المفردات من كل زوج بطريقة عشوائية ويتم إخضاعه لمعالجة معينة بينما تخضع المفردة الثانية في كل زوج لمعالجة أخرى حيث يكون الهدف في النهاية هو عقد مقارنة بين تأثيري المعالجتين على المفردات. ووجود التماثل والتشابه بين مفردتي كل زوج يخلق نوعا من الارتباط بين المشاهدات وبالتالي نحكم على هذه الحالة بعدم استقلال مجتمعي الدراسة. وللتوضيح نعطي فيما يلي أمثلة تطبيقية توضح مفهوم عدم الاستقلال.

المشاهدات قبل وبعد المعالجة:

افترض أننا نريد تقييم نظام غذائي معين لإتقاص الوزن. في هذه الحالة نعتبر أن أفضل أسلوب لإجراء هذه الدراسة هو أن نقوم باختيار مجموعة من الأفراد ممن يرغبون في اتباع هذا النظام ونسجل أوزانهم قبل بدء التجربة، ثم نترك كل فرد منهم يطبق هذا النظام لفترة زمنية محددة. ونقوم في نهاية التجربة بقياس الأوزان مرة أخرى ونأخذ حجم التغير في الوزن كمؤشر يعكس مدى فاعلية النظام الغذائي المقترح على تخفيض الوزن. في هذه الحالة نقول بأن بيانات أوزان جميع الأفراد الذين يمكن أن يطبقوا هذا النظام تمثل مجتمع، بينما تمثل بيانات أوزانهم بعد تطبيق النظام الغذائي

المجتمع الآخر. ولبيان أن مشاهدات المجتمعين مرتبطة ببعضها، افترض أننا بدأنا بفرد كان وزنه قبل تطبيق النظام هو ١٢٠ كجم. مهما كانت فاعلية النظام الغذائي نتوقع أن يظل وزنه بعد النظام مرتفعا أيضا وليكن ١٠٠ كجم على سبيل المثال. من جهة أخرى، إذا بدأنا بشخص وزنه ٨٠ كجم قد يصبح وزنه، على سبيل المثال، بعد تطبيق النظام ٧٥ كجم. أي أننا في مثل هذا النوع من التجارب نتوقع أن ترتبط القيم المرتفعة ببعضها البعض وأن ترتبط القيم المنخفضة ببعضها البعض. وبالتالي لا يوجد استقلال بين بيانات المجتمعين نتيجة أخذ القياسات على نفس الأفراد قبل وبعد التجربة.

أزواج الملاحظات المتماثلة

افترض أننا نريد المتارنة بين أسلوبيين مختلفين لتدريس مقرر معين على مستوى استيعاب الطلبة ودرجاتهم في امتحان لهذا المقرر. يقتضي إجراء هذه الدراسة أن يتم تدريس المقرر لمجموعة من الطلبة باستخدام الأسلوب الأول ولمجموعة أخرى باستخدام الأسلوب الثاني (لماذا لا نستخدم نفس المجموعة من الطلبة كما في الحالة السابقة؟). إذا قلنا بأننا سوف نقوم باختيار عينة عشوائية من الطلبة لكل أسلوب من أسلوبي التدريس وفي النهاية نعقد لهم امتحان موحد، قد نصل إلى نتائج مضللة. فعلى سبيل المثال، قد نجد أن درجات الطلبة الذين درسوا المقرر بالأسلوب الأول أفضل من تلك الخاصة بالذين درسوا المقرر بالأسلوب الثاني، ولكن هذا الاختلاف قد يكون راجعا إلى تفوق طلبة المجموعة الأولى على طلبة المجموعة الثانية، وليس نتيجة تأثير أسلوب التدريس الأول. وإذا قلنا أنه يجب أن يكون مستوى جميع الطلبة في المجموعتين متقارب، نجد أنه من الصعوبة

في العديد من الأحيان الحصول على عدد كبير من الطلبة ذوي المستويات المتقاربة لتقسيمهم عشوائيا على المجموعتين. في مثل هذه الحالات نلجأ، كما سبق أن ذكرنا، إلى تكوين أزواج من المفردات المتجانسة، فنقوم بتكوين مجموعة من الأزواج، بحيث يشتمل كل زوج على طالبين في نفس المرحلة الدراسية ولهما تقريبا نفس المستوى العلمي ونفس درجة الالتزام في حضور المحاضرات وما إلى غير ذلك من العوامل. بعد ذلك، نقوم باختيار أحد الطلبة من كل زوج بطريقة عشوائية ونرسله ليدرس المقرر بالأسلوب الأول ونرسل الآخر ليدرس المقرر بالأسلوب الثاني. وفي نهاية التجربة نعقد امتحان موحد لجميع الطلبة. ولبيان كيف أنه يوجد ارتباط بين مجموعتي البيانات، نقول بأنه إذا كان لدينا زوج المستوى العلمي للطلاب فيه مرتفع، فإتينا نتوقع أن يحصل كل منهما على درجات مرتفعة في الامتحان وذلك بصرف النظر عن طريقة التدريس. كذلك إذا كان لدينا زوج المستوى العلمي للطلاب فيه منخفض، نتوقع أن يحصل كل منهما على درجات منخفضة وذلك بصرف النظر عن أسلوب التدريس. أي أننا مرة أخرى نتوقع أن ترتبط المشاهدات المرتفعة ببعضها البعض والمنخفضة ببعضها البعض ويكون لدينا حالة عدم استقلال بين المجتمعين.

وترجع أهمية التمييز بين مجتمعي الدراسة من حيث الاستقلال أو عدم الاستقلال إلى أن أساليب الاستدلال الإحصائي المستخدمة تختلف في الحالتين. وسوف تشمل دراستنا في هذا الفصل المقارنة بين متوسطي مجتمعين (مستقلين أو غير مستقلين) والمقارنة بين نسبتي مجتمعين مستقلين وكذلك المقارنة بين تبايني مجتمعين مستقلين. وكما فعلنا في

الفصول السابقة، سوف نبدأ دائما بتحديد مقدرات النقاط، ثم توزيع المعايير لهذه المقدرات ثم نتناول أساليب اختبارات الفروض وتقدير فترات الثقة.

٤-٢ الاستدلال عن متوسطي مجتمعين مستقلين

نعلم من دراستنا السابقة أن مقاييس النزعة المركزية تعكس المستوى العام لقيم الظواهر التي تتم دراستها. فإذا أردنا عقد مقارنة بين مستوى القيم في مجتمعين مستقلين، فيمكن أن يتم ذلك بمقارنة متوسطي المجتمعين لتحديد ما إذا كانا متساويين، أم أن مستوى القيم في أحد المجتمعين أعلى منه في المجتمع الآخر. وعندما يتعذر استخدام أسلوب الحصر الشامل، تكون قيمة متوسط كل من المجتمعين مجهولة وبالتالي يتم تقديرها باستخدام متوسط عينة عشوائية مسحوبة من كل مجتمع وبعد ذلك نقوم بتطبيق مجموعة من أساليب الاستدلال الإحصائي للمقارنة بين المجتمعين مماثلة لما سبق وقلناه في الفصول السابقة.

وفي حالة دراسة مجتمعين أو أكثر تظهر الحاجة إلى استخدام أدلة سفلية تميز مقاييس المجتمع الأول عن مقاييس المجتمع الثاني كما يتضح من الجدول التالي.

المقياس	المجتمع الأول	المجتمع الثاني
متوسط المجتمع	μ_1	μ_2
الفرق بين متوسطي المجتمعين	$\mu_2 - \mu_1 = \Delta$ أو $\mu_1 - \mu_2 = \Delta$	
تباين المجتمع	σ_1^2	σ_2^2
التباين المشترك	σ^2	
متوسط العينة	\bar{x}_1	\bar{x}_2
حجم العينة	n_1	n_2
تباين العينة	s_1^2	s_2^2
التباين المشترك	$\frac{s_1^2(1-n_1) + s_2^2(1-n_2)}{n_1 + n_2 - 2} = s_p^2$	

٤-٢-١ مقدار نقطة للفرق بين متوسطي المجتمعين

استخدمنا من قبل متوسط العينة كمقدر نقطة لمتوسط المجتمع. وبالتالي يكون \bar{x}_1 مقدار نقطة لمتوسط المجتمع الأول μ_1 ، ويكون \bar{x}_2 مقدار نقطة لمتوسط المجتمع الثاني. ونستخدم في هذا الفصل الفرق بين متوسطي العينتين كمقدر نقطة للفرق بين متوسطي المجتمعين.

مقدر نقطة للفرق $\Delta = \mu_2 - \mu_1$ ، يكون الفرق بين متوسط العينة الأولى ومتوسط العينة الثانية $(\bar{x}_2 - \bar{x}_1)$. ومن ناحية أخرى، يكون

مقدر نقطة للفرق $\Delta = \mu_1 - \mu_2$ ، هو الفرق بين متوسط العينة الثانية والأولى $(\bar{S}_1 - \bar{S}_2)$. وحيث أن كلا من \bar{S}_1 و \bar{S}_2 يكون متغيرا عشوائيا، فإن الفرق بينهما يكون هو الآخر متغيرا عشوائيا، ويكون له توزيع احتمالي هو توزيع المعاينة للفرق بين متوسطي عينتين.

$$4-2-2 \text{ توزيع المعاينة للفرق } \bar{S}_1 - \bar{S}_2$$

إذا فرضنا أن مجتمعي الدراسة مستقلان ولكل منهما توزيع طبيعي، ع (μ_1, σ_1^2) للمجتمع الأول و ع (μ_2, σ_2^2) للمجتمع الثاني، فمن نظرية 1-1 بالفصل الأول، يكون لمتوسط عينة المجتمع الأول، \bar{S}_1 ، توزيع طبيعي ع $(\mu_1, \frac{\sigma_1^2}{n_1})$. كذلك، يكون لمتوسط عينة المجتمع الثاني، \bar{S}_2 ،

توزيع طبيعي ع $(\mu_2, \frac{\sigma_2^2}{n_2})$. وحيث أن مجتمعي الدراسة مستقلان، فإن

المتغيرين العشوائيين يكونان مستقلان أيضا. وإذا استخدمنا الرمز $\hat{\Delta}$ للتعبير عن مقدر النقطة للفرق بين متوسط المجتمع الأول والثاني، أي أن:

$$\hat{\Delta} = \bar{S}_1 - \bar{S}_2$$

فإن $\hat{\Delta}$ يكون دالة خطية في متغيرين مستقلين ولكل منهما توزيع طبيعي، وبالتالي يكون للمتغير العشوائي $\hat{\Delta}$ توزيع طبيعي أيضا. ولتحديد متوسط وتباين هذا التوزيع لاحظ أن:

$$1 = 1 \quad \text{و} \quad 1 = 2$$

بالتالي يكون الوسط الحسابي لتوزيع $\hat{\Delta} = \bar{y}_1 - \bar{y}_2$ من نظرية 1-1 هو:

$$y_1 \mu (1) + y_2 \mu (2) = \bar{y}_1 \mu + \bar{y}_2 \mu$$

$$y_1 \mu - y_2 \mu =$$

ويكون تباين توزيع $\hat{\Delta}$ هو:

$$\frac{y_1^2 \sigma^2}{n_1} (1) + \frac{y_2^2 \sigma^2}{n_2} (2) = \bar{y}_1^2 \sigma^2 + \bar{y}_2^2 \sigma^2$$

$$= \frac{y_1^2 \sigma^2}{n_1} + \frac{y_2^2 \sigma^2}{n_2}$$

لاحظ أن تباين الفرق بين متوسطي العينتين هو مجموع تباينيهما. وذلك لأن التباين صيغة تربيعية تتحول فيها الإشارة السالبة إلى إشارة موجبة. ونخلص من هذا أنه عندما يكون مجتمع الدراسة مستقلين ولكل منهما توزيع طبيعي، يكون للفرق بين متوسطي العينتين المسحوبتين منهما توزيع طبيعي

$$\left(\frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \right) \sim N(0, 1)$$

ولحساب احتمالات أية صيغ للفرق بين متوسطي العينتين، نستخدم الوحدات المعيارية:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

حيث يكون للمتغير Z توزيع طبيعي معياري ع (صفر ، 1).

مثال ٢-١

سحبت عينة عشوائية حجمها ٢٥ مفردة من مجتمع له توزيع طبيعي ع (٣٥ ، ١٦)، وسحبت عينة عشوائية حجمها ٥٠ مفردات من مجتمع طبيعي آخر ع (٣٥ ، ٨) ومستقل عن المجتمع الأول. ما هو احتمال ألا يتجاوز الفرق بين متوسطي العينتين وحدتان فقط لا غير؟

الحل:

إذا فرضنا أن متوسط عينة المجتمع الأول هو \bar{X}_1 ، وأن متوسط عينة المجتمع الثاني هو \bar{X}_2 ، يكون المطلوب هو حساب احتمال الصيغة
 $(|\bar{X}_1 - \bar{X}_2| \leq 2)$. وهنا أخذنا الفرق المطلق بين متوسطي العينتين لأن المطلوب لم يتضمن أي اتجاه للفرق بينهما.

$$\begin{aligned} & P(|\bar{X}_1 - \bar{X}_2| \leq 2) \\ &= P(-2 \leq \bar{X}_1 - \bar{X}_2 \leq 2) \\ & \text{من المعطيات نجد أن:} \end{aligned}$$

$$\mu_2 - \mu_1 = 37 - 37 = \text{صفر}.$$

$$\frac{1,0}{50} + \frac{16}{25} = \frac{2\sigma}{20} + \frac{2\sigma}{10}$$

$$= 0,81.$$

وبتحويل حدود المتباينة في صيغة الاحتمال السابق إلى وحدات معيارية تصبح على الصورة:

$$ح(2 \geq \bar{X}_1 - \bar{X}_2 \geq 2)$$

=

$$\left(\frac{2 - \text{صفر}}{0,81\sqrt{}} \geq \frac{-(\bar{X}_1 - \bar{X}_2) - \text{صفر}}{0,81\sqrt{}} \geq \frac{2 - \text{صفر}}{0,81\sqrt{}} \right) ح$$

$$ح(2,22 \geq Z \geq -2,22) =$$

$$= \Phi(2,22) - \Phi(-2,22)$$

$$= 1 - \Phi(2,22)$$

$$= 1 - 0,9868 \times 2 =$$

$$= 0,0272.$$

نخلص من هذا المثال إلى أنه إذا كان لمجمعي الدراسة نفس المتوسط، فإننا نتوقع أن يكون الفرق بين متوسطي العينتين طفيفا باحتمالات كبيرة. وهذا يتضمن أن الفرق بين متوسطي العينتين يعكس بصورة جيدة حقيقة الفرق بين متوسطي المجتمعين.

وجدنا في الفصل الاول أن حد خطأ التقدير باحتمال $1 - \alpha$ لتقدير متوسط المجتمع μ يكتب على الصورة

$$Z = \delta \times \frac{\alpha}{\sqrt{2}}$$

وإذا كتبنا خطأ تقدير الفرق بين متوسطي مجتمعين باستخدام الفرق بين متوسطي عينتين مسحوبتين من هذين المجتمعين على الصورة:

$$\text{خطأ التقدير} = |(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)|$$

فيمكن استخدام علاقة مماثلة لإيجاد حد خطأ تقدير الفرق بين متوسطي المجتمعين باستخدام الفرق بين متوسطي عينتين مسحوبتين من المجتمعين باحتمال $1 - \alpha$. وتكون صيغة خطأ التقدير في هذه الحالة، ومن توزيع المعاينة للفرق بين متوسطي العينتين، على الصورة:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

مثال ٤-٢

تم اختيار عينة عشوائية حجمها ٢٠ مفردة من مجتمع له توزيع طبيعي ع) (١٦ ، ١٤) ، وعينة عشوائية حجمها ٢٥ مفردة من مجتمع له توزيع طبيعي ع) (١٠,٢٥ ، ٢,٤) .

المطلوب:

- 1- ما هو حد خطأ تقدير الفرق بين متوسطي المجتمعين باستخدام الفرق بين متوسطي العينتين عند احتمال ٩٥ % ؟
- 2- إذا أردنا اختيار عيتين من المجتمعين لهما نفس الحجم، ن، ما هي قيمة ن التي تجعل أقصى حجم لخطأ تقدير الفرق بين متوسطي المجتمعين وحدة واحدة ونصف باحتمال ٩٩ %.

الحل:

1- حد خطأ التقدير باحتمال 95%

العينة الأولى : $n_1 = 20$ $\sigma_1^2 = 16$

العينة الثانية : $n_2 = 25$ $\sigma_2^2 = 10,25$

$$-1 = \alpha - 0,95 \quad \alpha = 0,05 \quad -1 = \alpha - 0,95$$

$$Z_{1-\alpha/2} = 1,96$$

$$\delta = 1,96 \times \sqrt{\frac{16}{20} + \frac{10,25}{25}}$$

$$\delta = 1,96 \times \sqrt{1,21}$$

$$\delta = 2,156$$

2- تحديد حجم العيّنتين المتساوي

$$2,076 = 0.995Z = 2\alpha-1Z \quad 0.99 = \alpha-1 \quad 1,0 = \delta$$

$$\sqrt{\frac{10,25}{n} + \frac{16}{n}} \times 2,076 = 1,0$$

$$\sqrt{\frac{26,25}{n}} \times 2,076 = 1,0$$

إن

$$1,798 = \frac{0,123 \times 2,076}{1,0} = \sqrt{\frac{\quad}{n}}$$

وبالتالي يكون حجم العينة التي يجب اختيارها من كل مجتمع هو:

$$n = 78 \text{ مفردة تقريبا.}$$

٤-٢-٣ تقدير فترات الثقة واختبارات الفروض للفرق بين

متوسطي مجتمعين طبيعيين معلومي التباين

لإيجاد مقدر فترة ثقة $(1-\alpha)\%$ للفرق بين متوسطي مجتمعين لهما توزيعان طبيعيان تباينهما معلوم، نستخدم نفس التعريف الذي قدمناه عند تقدير فترة ثقة لمتوسط مجتمع، حيث نكرنا أن حدود فترة الثقة تحسب بطرح وإضافة حد خطأ التقدير عند احتمال $1 - \alpha$ إلى مقدر النقطة.

وبالتالي تكتب حدود فترة ثقة $(1-\alpha)\%$ للفرق بين متوسطي المجتمعين الأول والثاني $\mu_1 - \mu_2$ على الصورة:

$$\frac{\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}{(1-\alpha)\%}$$

وإذا أردنا إيجاد مقدار فترة ثقة للفرق بين متوسطي المجتمعين الثاني والأول $\mu_2 - \mu_1$ ، فإننا نعكس اتجاه الطرح في صيغة فترة الثقة السابقة ليكون $(\bar{X}_2 - \bar{X}_1)$.

لإجراء اختبارات فروض تتعلق بالفرق بين متوسطي مجتمعين، فإننا نتناول المراحل الأربع السابق تقديمها في الفصل الثالث، ولكن في حالتنا الراهنة تكون معلمة الاختبار هي $\Delta = \mu_1 - \mu_2$ ، ولكننا $(\Delta = \mu_2 - \mu_1)$. ويمكن إجراء الاختبارات لأية قيمة نظرية Δ_0 ، ولكننا سوف نقتصر على الحالة الأكثر شيوعاً في الحياة العملية والتي تتضمن وجود أو عدم وجود فرق بين متوسطي المجتمعين. أي أننا سوف نقتصر على دراسة الحالة $\Delta_0 = \text{صفر}$.

المرحلة الأولى: الفروض الإحصائية

في مشاكل اختبار معنوية الفرق بين متوسطي مجتمعين، تكون الفروض الإحصائية على أحد الأشكال الثلاثة السابق دراستها في اختبارات الطرفين، اختبار الطرف الأيمن واختبار الطرف الأيسر.

اختبار الطرفين:

يستخدم اختبار الطرفين في الحالة التي نريد فيها اختبار ما إذا كان للمجتمعين نفس المتوسط أم يوجد بينهما اختلاف، وذلك بصرف النظر عن اتجاه هذا الاختلاف. فعلى سبيل المثال، إذا أردنا اختبار الفرض بأنه لكل من الطلاب والطالبات نفس مستوى الأداء في امتحان مقرر معين في مقابل الفرض بأن إحدى المجموعتين أفضل من الأخرى، دون تحديد اتجاه الأفضلية، فإننا نستخدم اختبار الطرفين. ويمكن أن تكتب الفروض الإحصائية في هذه الحالة على إحدى الصور:

$$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0 \quad \text{مقابل}$$

أو

$$H_0: \mu = \mu_0 \quad H_1: \mu < \mu_0 \quad \text{ضد}$$

أو:

$$H_0: \mu = \mu_0 \quad H_1: \mu > \mu_0 \quad \text{ضد}$$

اختبار الطرف الأيمن:

يستخدم هذا الاختبار إذا كان القرار الذي نرغب في الوصول إليه يتعلق بتحديد ما إذا كان متوسط المجتمع الأول (μ_1) يزيد بصورة جوهرية عن متوسط المجتمع الثاني (μ_2) أم لا (يساويه أو قد يقل عنه). على سبيل المثال، عند مقارنة مستوى أداء كل من الطلاب والطالبات في امتحان معين، إذا كان متوسط درجات الطلاب هو μ_1 ، ومتوسط درجات الطالبات هو μ_2 ، وأردنا اختبار ما إذا كان الطلاب أفضل من الطالبات، تكون الفروض الإحصائية على إحدى الصور:

$$\mu_1 \geq \mu_2 \quad H_0 \quad \text{مقابل} \quad \mu_1 < \mu_2 \quad H_1$$

أو

$$\mu_1 - \mu_2 \geq \text{صفر} \quad H_0 \quad \text{مقابل} \quad \mu_1 - \mu_2 < \text{صفر} \quad H_1$$

أو:

$$\Delta > \text{صفر} \quad H_0 \quad \text{مقابل} \quad \Delta \leq \text{صفر} \quad H_1$$

$$(\text{حيث } \Delta = \mu_1 - \mu_2)$$

لاحظ أن تعريف Δ يعتبر أساسي في هذه الحالة وذلك لأننا إذا ما عرفنا Δ بصورة عكسية، أي $\Delta = \mu_2 - \mu_1$ ، فإن وجهة الاختبار سوف تتغير ويصبح اختبار طرف أيسر بدلا من اختبار طرف أيمن.

اختبار الطرف الأيسر:

يستخدم هذا الاختبار إذا كان القرار الذي نرغب في الوصول إليه يتعلق بتحديد ما إذا كان متوسط المجتمع الأول (μ_1) يقل بصورة جوهرية عن متوسط المجتمع الثاني (μ_2) أم لا (يساويه أو قد يزيد عنه). على سبيل المثال، عند مقارنة مستوى أداء كل من الطلاب والطالبات في امتحان معير إذا كان متوسط درجات الطلاب هو μ_1 ، ومتوسط درجات الطالبات هو μ_2 ، وأردنا اختبار ما إذا كن الطالبات أفضل من الطلاب أم لا، تكون الفرضية الإحصائية على إحدى الصور:

$$\mu_1 > \mu_2 : H_0 \quad \text{مقابل} \quad \mu_1 \leq \mu_2 : H_1$$

أو

$$\mu_1 - \mu_2 \leq \text{صفر} : H_0 \quad \text{مقابل} \quad \mu_1 - \mu_2 > \text{صفر} : H_1$$

أو:

$$\Delta \leq \text{صفر} : H_0 \quad \text{مقابل} \quad \Delta > \text{صفر} : H_1$$

$$(\text{حيث } \Delta = \mu_1 - \mu_2)$$

المرحلة الثانية: إحصائية الاختبار

سبق أن وجدنا من توزيع المعاينة للفرق بين متوسطي عینتين مسحوبتين من توزيعين طبيعيين ومستقلين أن المتغير العشوائي

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

يكون له توزيع طبيعي معياري. وإذا ما استخدمنا الصيغة السابقة لإحصائية للاختبار، فإننا نضع الفرق $\mu_1 - \mu_2$ مساويا للصفر وذلك لأننا وكما ذكرنا من قبل نجري اختبارات الفروض في ظل صحة فرض العدم. وبالتالي تصبح إحصائية الاختبار على الصورة:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

ويلاحظ أن البسط في صيغة Z^* يشتمل على الفرق بين متوسطي العينتين. وكلما قل الفرق بين المتوسطين، وبالتالي قلت القيمة العددية لإحصائية الاختبار واقتربت من الصفر، كلما كان لدينا دلالة أقوى على عدم وجود فرق بين متوسطي المجتمعين وعلى صحة فرض العدم. ومن ناحية أخرى، كلما زاد الفرق بين متوسطي العينتين وبالتالي زادت القيمة العددية لإحصائية الاختبار، كلما زادت الدلالة على عدم صحة فرض العدم.

المرحلة الثالثة: تحديد القيمة الحرجة

حيث أننا في هذه المرحلة نفترض أن تباين كل من المجتمعين معلوم، فإننا نقوم بتحديد القيمة الحرجة من جداول التوزيع الطبيعي المعياري. وهي كما سبق $Z_{1-\alpha}$ في حالة اختبار الطرف الأيمن، $-Z_{1-\alpha}$ في حالة اختبار الطرف الأيسر، و $\pm Z_{1-\alpha/2}$ في حالة اختبار الطرفين.

المرحلة الرابعة: تقرير نتيجة الاختبار

إذا وقعت قيمة إحصائية الاختبار داخل المنطقة الحرجة، أي إذا زادت القيمة العددية لإحصائية الاختبار عن القيمة الجدولية، فإننا نرفض فرض عدم ونقبل الفرض البديل ونستنتج وجود اختلاف بين متوسطي المجتمعين أو أن متوسط أحدهما يكون أكبر من متوسط الآخر. كذلك يمكن تقرير نتيجة الاختبار بحساب القيمة الاحتمالية بنفس الأسلوب السابق شرحه في الفصل السابق.

مثال ٤-٣

في عينة من ٢٠٠ عامل من العاملين بشركات المقاولات والإنشاءات وجد أن المتوسط الشهري لأجر العامل هو ٥٣٨ جنيها. وفي عينة من ١٥٠ عاملا بالشركات الصناعية وجد أن المتوسط الشهري لأجر العامل هو ٥١٢ جنيها. وعلى فرض أن توزيع أجور العمال في كل من شركات القطاعين هو توزيع طبيعي باحتراف معياري ٦٦ و ٦٠ جنيها على الترتيب، فالمطلوب:

- ١- ما هو مقدر نقطة للفرق بين متوسطي أجور العمال في قطاع المقاولات وقطاع الصناعة؟
- ٢- أوجد تقدير نقطة للفرق بين متوسطي الأجور في قطاعي المقاولات والصناعة.
- ٣- أوجد تقدير فترة ثقة ٩٥% للفرق بين متوسطي الأجور في المطلوب السابق.
- ٤- اختبار الفرض بأن مستوى أجور العمال يزيد بصورة جوهرية في قطاع المقاولات عنه في قطاع الصناعة وذلك عند مستوى معنوية ٥%.

الحل:

يلاحظ أن لكل قطاع طبيعة خاصة تحدد مستويات أجوره، وبالتالي نعتبر أن مجتمعي الدراسة مستقلين عن بعضهما البعض.

صناعة	مقاولات	
μ_1	μ_2	متوسط المجتمع
σ_1	σ_2	الانحراف المعياري
$n_1 = 360$	$n_2 = 430$	تباين المجتمع
$\bar{x}_1 = 512$	$\bar{x}_2 = 538$	متوسط العينة
$n_1 = 150$	$n_2 = 200$	حجم العينة

1- من الفصل الأول يكون مقدر النقطة هو صيغة إحصاء العينة

المستخدمة لتقدير المعطمة. وبالتالي يكون مقدر نقطة للفرق بين متوسطي المجتمعين هو الفرق بين متوسطي العينتين. ومن السؤال نجد أن المطلوب

هو مقدر نقطة للفرق $\mu_2 - \mu_1$

$$\overline{S_2} - \overline{S_1} = \hat{\mu}_2 - \hat{\mu}_1 = 211 - 111$$

2- من الفصل الأول يكون تقدير النقطة هو قيمة المقدر محسوبة من بيانات العينة المتاحة. وبالتالي يكون تقدير النقطة للفرق بين متوسطي الأجور حسب المطلوب هو:

$$\hat{\mu}_2 - \hat{\mu}_1 = 538 - 512 = 26 \text{ جنيها.}$$

3- تقدير فترة ثقة للفرق بين متوسطي المجتمعين
حيث أن توزيعي المجتمعين هما توزيعان طبيعيان معلوما التباين، ومجمعي الدراسة مستقلان، فإننا نقوم بتقدير فترة الثقة وإجراء اختبار الفروض في المطلوب التالي باستخدام التوزيع الطبيعي المعياري.

$$1.96 = 0.975Z = 2\alpha-1Z \quad 0.05 = \alpha \quad 0.95 = 1 - \alpha$$

$$\sqrt{\frac{3600}{100} + \frac{4306}{200}} \quad 1,96 \pm 26$$

$$٦,٧٧ \times ١,٩٦ \pm ٢٦$$

$$١٣,٢٦ \pm ٢٦$$

الحد الأدنى للفرق بين المتوسطين = ١٢,٧٤ جنيها.

الحد الأعلى للفرق بين المتوسطين = ٣٩,٢٦ جنيها.

4- قرصّ البحث الذي نريد اختياره يتعلق بما إذا كان متوسط أجور العمال في قطاع المقاولات، μ_1 ، يزيد جوهريا عنه في قطاع الصناعة μ_2 ، أم لايزيد (يساوي أو قد يقل). وبالتالي يكون الاختبار في هذه الحالة هو اختبار طرف أيمن.

الفروض الإحصائية:

$$H_0: \mu_2 \geq \mu_1 \quad \text{مقابل} \quad H_1: \mu_2 < \mu_1$$

إحصائية الاختبار

$$Z^* = \frac{(\bar{x}_2 - \bar{x}_1)}{\sqrt{\frac{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}}}$$

$$Z^* = \frac{٥١٢ - ٥٣٨}{\sqrt{\frac{٣٦٠٠}{١٥٠} + \frac{٤٣٥٦}{٢٠٠}}}$$

$$٣,٨٤ = \frac{٢٦}{٦,٧٧} =$$

القيمة الحرجة

$$\alpha = 0.05 - \text{اختبار طرف أيمن}$$

$$1.645 = 0.95Z = \alpha_1 Z$$

نتيجة الاختبار

حيث أن قيمة إحصائية الاختبار تزيد عن القيمة الحرجة، فإننا نرفض فرض العدم ونقبل الفرض البديل ونستنتج أن متوسط أجور العمال في قطاع المقاولات يزيد عن متوسط أجور العمال في قطاع الصناعة وذلك عند مستوى معنوية ٥%.

٤-٢-٤ تقدير فترات الثقة واختبارات الفروض للفرق بين متوسطي مجتمعين غير معلومي التباين.

سبق أن رأينا في الفصلين الثاني والثالث عند الاستدلال عن متوسط مجتمع غير معلوم التباين أننا نقوم باستخدام تباين العينة، ع^٢ كمقدر نقطة لتباين المجتمع. ويكون توزيع المعاينة لإحصائية الاختبار عندئذ هو توزيع t وليس التوزيع الطبيعي المعياري. وفي الفصل الحالي أيضا عند الاستدلال عن الفرق بين متوسطي مجتمعين لهما توزيع طبيعي، إذا كان تباين المجتمعين غير معلوم، فإننا سوف نقوم باستخدام توزيع t . وسوف نفرق في هذا الصدد بين حالتين هما، الحالة التي يكون فيها

للمجتمعين تباين متساوي، أي $\sigma^2_1 = \sigma^2_2 = \sigma^2$ ، والحالة التي يكون فيها

للمجتمعين تباين غير متساوي، أي $\sigma^2_1 \neq \sigma^2_2$.

الحالة الأولى: تباين المجتمعين متساوي

إذا أمكننا افتراض أن لمجمعتي الدراسة نفس التباين، سواء كان هذا الافتراض معلوما نظريا أو تم اختباره، كما سنرى في نهاية هذا الفصل، تكتب صيغة Z^* من توزيع المعاينة للفرق بين متوسطي مجتمعين على الصورة:

$$Z^* = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

وإذا كانت قيمة σ^2 غير معلومة، فإننا نقوم بتقديرها من البيانات المتاحة. في هذه الحالة، يعتبر تباين العينة الأولى، s^2_1 ، مقدر نقطة لتباين المجتمع σ^2_1 . وكذلك يكون تباين العينة الثانية، s^2_2 ، مقدر نقطة لتباين المجتمع. وحيث أن تباين المجتمعين متساوي، فإننا نقوم باستخدام متوسط مرجع لتباين العينتين معا (حيث يكون الترجيح على أساس درجات الحرية لكل عينة)، لتقدير σ^2 . ويطلق على مقدر تباين المجتمعين المتساوي اسم التباين المشترك ونرمز له بالرمز s^2_m ويحسب باستخدام العلاقة:

$$\bar{E} = \frac{\sum_{i=1}^n (E_i - \bar{E})^2}{n-1}$$

ويمكن أيضا استخدام الصيغة المكافئة التالية

$$\bar{E} = \frac{\sum_{i=1}^n (E_i^2) - \frac{(\sum_{i=1}^n E_i)^2}{n}}{n-1}$$

وعند استبدال σ^2 في صيغة Z^* السابقة بالمقدر \bar{E} ، فإننا نحصل على متغير عشوائي جديد هو t^* له توزيع t بدرجات حرية $(n_1 + n_2 - 2)$ ونستخدمه كما ذكرنا في بناء فترات الثقة وإجراء اختبارات الفروض للفرق بين متوسطي المجتمعين.

ويكون مقدر فترة ثقة $(1 - \alpha)\%$ للفرق بين متوسطي المجتمعين الأول والثاني $\mu_1 - \mu_2$ على الصورة:

$$\left(\bar{E} \pm t_{\alpha/2, n-2} \sqrt{\frac{\bar{E}}{n-2}} \right)$$

وفي اختبارات الفروض، تبقى الفروض الإحصائية كما هي على أحد الصور الثلاث السابق تقديمها، وتصبح إحصائية الاختبار على الصورة:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

وفي المرحلة الثالثة، يتم تحديد القيمة الحرجة من جداول توزيع t بدرجات حرية $(n_1 + n_2 - 2)$ وعند الاحتمال المناسب حسب نوع الاختبار. ويتم تقرير نتيجة الاختبار كالمعتاد بمقارنة قيمة إحصائية الاختبار بالقيمة الحرجة كما سبق بيانه.

الحالة الثانية: تباين المجتمعين غير متساوي

إذا كانت هناك دلالة قوية على عدم تساوي تبايني المجتمعين، فإنه لا يصح في هذه الحالة استخدام التباين المشترك، وإنما نقوم بتقدير تباين كل مجتمع على حدة باستخدام تباين العينة الخاصة به. وتصبح صيغة كل من مقدار فترة الثقة وإحصائية الاختبار للفرق $\mu_1 - \mu_2$ على الترتيب على الصورة:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

ويتم الكشف في جداول توزيع t عند درجات حرية d ، ومستوى المعنوية المطلوب. وتحدد درجات الحرية d بأقرب عدد صحيح أصغر من قيمة المقدار:

$$d = \frac{\left(\frac{x_1}{n_1} + \frac{x_2}{n_2} \right)}{\left(\frac{x_1}{n_1} + \frac{x_2}{n_2} \right)} = \frac{\left(\frac{x_1}{n_1} + \frac{x_2}{n_2} \right)}{\left(\frac{x_1}{n_1} + \frac{x_2}{n_2} \right)}$$

وفي الحالتين السابقتين عندما يكون حجم كل من العينتين كبيراً بدرجة كافية، فإننا نعود إلى استخدام التوزيع الطبيعي المعياري كتقريب لتوزيع t للعينات كبيرة الحجم.

مثال ٤-٤

للمقارنة بين مستوى كفاءة الذكور والإناث في استخدام معالج كلمات النوافذ (WINWORD)، طلب من كل فرد في عينة عشوائية من عشرة ذكور وعينة عشوائية من ثماني إناث، ممن لهم تقريبا نفس مستوى

التدريب، كتابة مقال معين باستخدام هذا البرنامج. وفي نهاية التجربة تم قياس الزمن الذي استغرقه كل منهم في الكتابة فكان متوسط الزمن المستغرق في عينة الذكور ٦,٤ دقيقة بالاحراف معياري ٦٠ ثانية، وكان متوسط الزمن في عينة الإناث ٥,٢ دقيقة بالاحراف معياري ٤٨ ثانية. وعلى فرض أن تباين زمن الكتابة متساوي في مجتمعي الدراسة، وأن كلا منهما يتبع توزيعاً طبيعياً تقريباً، فالمطلوب:

- 1- هل تؤيد هذه البيانات القول بأن الإناث أفضل من الذكور في استخدام البرنامج عند مستوى معنوية 1%.
- 2- أوجد تقدير فترة ثقة 99% للفرق بين متوسط زمن الذكور ومتوسط زمن الإناث.

الحل:

يلاحظ في هذه المشكلة أن مجتمعي الدراسة مستقلين عن بعضهما البعض وذلك لأن الزمن الذي يستغرقه أي فرد في عينة الذكور لا تربطه أية صلة بالزمن الذي تستغرقه أية مفردة في عينة الإناث. كذلك يجب ملاحظة أن متوسط الزمن في البيانات تم قياسه بالدقائق، بينما تم قياس الانحراف المعياري للزمن بالثواني. في هذه الحالة يجب توحيد وحدات القياس، فإما أن نحول المتوسط إلى ثواني، أو نحول الانحراف المعياري إلى دقائق.

إناث	ذكور	
μ_2	μ_1	متوسط الزمن في المجتمع
σ_2^2	σ_1^2	تباين المجتمع (متساوي)
$\bar{y}_2 = 0,2$	$\bar{y}_1 = 6,4$	متوسط العينة
$s_2^2 = 0,8$	$s_1^2 = 1$	الانحراف المعياري
$y_2 = 0,64$	$y_1 = 1$	تباين العينة
$n_2 = 8$	$n_1 = 10$	حجم العينة

حيث أن تباين المجتمعين متساوي، نقوم بحساب التباين المشترك

$$s_p^2 = \frac{(0.64)(1-8) + (1)(1-10)}{2-8+10} = 0,8420$$

حيث أن العينتين مسحوبتان من مجتمعين طبيعيين ومستقلين وحجم كل من العينتين صغير، فإننا نستخدم توزيع t لتقدير فترة الثقة وإجراء اختبار الفروض.

1- هل الإناث أفضل من الذكور؟

تقتضي الإجابة على هذا السؤال إجراء اختبار للفروض تتم فيه صياغة فرض العدم على أساس أن الإناث ليسوا أفضل من الذكور (لا يوجد فرق في الكفاءة أو على العكس قد يكون الذكور أفضل). من جهة أخرى يتم وضع الفرض البديل على أساس أن الإناث أفضل. وحيث أن معيار الكفاءة

هنا هو استغراق زمن أقل، فإن الاختبار يكون ذو طرف أيمن إذا ما تم طرح متوسط الإناث من متوسط الذكور.

الفروض الإحصائية:

$$H_0: \mu_1 - \mu_2 \geq 0 \quad \text{صفر} \quad H_1: \mu_1 - \mu_2 < 0 \quad \text{صفر}$$

إحصائية الاختبار

$$t = \frac{0.2 - 6.4}{\sqrt{\left(\frac{1}{8} + \frac{1}{10}\right) \cdot 0.8425}} = -2.706$$

القيمة الحرجة

$$\alpha = 0.01 \quad d.f. = 16 + 10 - 2 = 16$$

$$t_{(0.01, 16)} = -2.083$$

نتيجة الاختبار

حيث أن قيمة إحصائية الاختبار تزيد عن القيمة الحرجة، فإننا نرفض فرض العدم ونقبل الفرض البديل ونستنتج أن الإناث أكثر كفاءة من الذكور وذلك عند مستوى معنوية 1%.

يلاحظ في هذا المثال أن قيمة إحصائية الاختبار قريبة من القيمة الحرجة. وفي الحياة العملية، يفضل في مثل هذه الحالات استخدام عينة أخرى أو زيادة حجم العينة حتى تكون لدينا دلالة كافية لصالح أحد الفرضين.

2- تقدير فترة ثقة 99% للفرق بين متوسطي زمن الذكور والإناث

$$0.01 = \alpha \quad 0.005 = 2\alpha \quad d = 16$$

$$t_{(0.005, 16)} = 2.924$$

$$\left(\frac{1}{n_1} + \frac{1}{n_2} \right)^2 \times (21\alpha, 2 - 21\alpha, n_1, n_2) \pm (\bar{y}_2 - \bar{y}_1)$$

$$\left(\frac{1}{8} + \frac{1}{10} \right)^2 \times 0.8425 \times 2.924 \pm (5.2 - 6.4)$$

$$0.4354 \times 2.924 \pm 1.2$$

$$1.27 \pm 1.2$$

الحد الأدنى = $1.27 - 1.2 = 0.07$ دقيقة.

الحد الأعلى = $1.27 + 1.2 = 2.47$ دقيقة.

مثال ٤-٥

أراد مدير شركة معينة دراسة تأثير نظامين للأجور على إنتاجية العمال. يتضمن النظام الأول إعطاء أجر شهري ثابت للعمال، بينما يتضمن النظام الثاني إعطاء العامل أجر متغير يرتبط بمعدلات الإنتاج. ولإجراء هذه الدراسة، قام المدير بتطبيق النظام الأول على عينة من ٣٢ عامل يعملون في أحد فروع الشركة، فوجد أن متوسط الإنتاج الشهري للعامل هو ٢٣٥ وحدة بانحراف معياري ٨ وحدات. كذلك قام بتطبيق النظام الثاني على

عينة من ٥٠ عامل في فرع آخر من فروع الشركة، فوجد أن متوسط الإنتاج الشهري للعامل هو ٢٦٠ وحدة بالتحرقا معياري ٢٠ وحدة. نتاجا على ما تقدم، قرر المدير تطبيق نظام الأجر المتغير على كافة العاملين في جميع فروع الشركة.

- 1- هل تتفق معه في هذا القرار عند مستوى معنوية ٥% ؟
- 2- أوجد تقدير فترة ثقة 95% للفرق بين متوسطي الإنتاجية باستخدام اختبار

الأجر المتغير والأجر الثابت.

الحل:

يلاحظ أن حجمي العينتين كبير في هذا المثال، وبالتالي نقوم بإجراء اختبار الفروقات بين متوسطي الإنتاجية باستخدام توزيع Z، وذلك على أساس من أن تباين المجتمعين مجهول. كذلك لم ينص المثال على تساوي التباينين، وبالتالي لا نقوم بحساب التباين المشترك وإنما نقوم بالتقدير من مجتمع بتباين العينة المسحوبة ٢٠٠ من جهة ثالثة، حيث أن الدراسة قد تمت في فرعين مختلفين للشركة وطبق النظام على مجاوي مختلفتين من العمال، تكون البيانات مستقلة عن بعضها البعض. العنيتين.

1- هل نظام الأجر المتغير أفضل؟

إن قرار مدير الشركة بتصميم نظام الأجر المتغير على جميع العاملين يعني أنه أفضل من حيث أنه يعطي مستوى إنتاجية أعلى. ونقوم نحن بإجراء الاختبار لتحديد ما إذا كانت الزيادة في الإنتاجية في ظل نظام الأجر المتغير

على نظيرتها في ظل نظام الأجر الثابت هي زيادة جوهرية، أم تعزى إلى عوامل الصدفة ولا يوجد لاختلاف بين النظامين.

أجر ثابت	أجر متغير	
μ_2	μ_1	متوسط المجتمع
$\sigma^2 = 235$	$\sigma^2 = 260$	متوسط العينة
$s^2 = 28$	$s^2 = 20$	الاحراف المعياري
$\sigma^2 = 26$	$\sigma^2 = 40$	تباين العينة
$n = 22$	$n = 50$	حجم العينة

الفروض الإحصائية

يكون الاختبار وفقا لتعريفنا لبيانات الأجر المتغير على أنه يمثل بيانات المجتمع الأول، هو اختبار طرف أيمن، وذلك لأننا نضع فرض العدم على أساس أن مستوى الإنتاجية في ظل الأجر المتغير ليس أفضل (يكون مثل المتوسط في ظل الأجر الثابت وقد يكون أسوأ). من ناحية أخرى، يصاغ الفرض البديل على أن متوسط الإنتاجية في ظل نظام الأجر المتغير هو الأكبر. بالتالي تكون الفروض الإحصائية على الصورة:

$$H_0: \mu_2 \geq \mu_1 \quad \text{مقابل} \quad H_1: \mu_2 < \mu_1$$

أو

$$H_0: \mu_2 - \mu_1 \geq \text{صفر} \quad \text{مقابل} \quad H_1: \mu_2 - \mu_1 < \text{صفر}$$

إحصائية الاختبار

$$Z^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$7.9 = \frac{235 - 260}{\sqrt{\frac{64}{32} + \frac{400}{50}}}$$

القيمة الحرجة

$$\alpha = 0.05 \quad \text{اختبار طرف أيمن} \quad Z_{1-\alpha} = Z_{0.95} = 1.645$$

نتيجة الاختبار

حيث أن إحصائية الاختبار تزيد كثيرا عن القيمة الحرجة، نرفض فرض العدم ونقبل الفرض البديل ونستنتج أن متوسط الإنتاجية في ظل نظام الأجر المتغير يزيد جوهريا عنه في ظل نظام الأجر الثابت وذلك عند مستوى معنوية 5%. وبالتالي نتفق مع مدير الشركة فيما توصل إليه من قرار.

القيمة الاحتمالية:

$$P(Z < 7.9) = 1 - P(Z > 7.9) = 1 - 0 = 1$$

- تقدير فترة ثقة 95% للفرق بين متوسطي الإنتاجية

$$\frac{\sqrt{\frac{24}{20} + \frac{24}{10}}}{1.96 \pm (s_2 - s_1)}$$

$$1.0 \sqrt{20} \times 1.96 \pm 6.2$$

$$\text{الحد الأدنى لفترة الثقة} = 6.2 - 20 = 13.8$$

$$\text{الحد الأعلى لفترة الثقة} = 6.2 + 20 = 26.2$$

٣-٤ الاستدلال عن متوسطي مجتمعين غير مستقلين

تناولنا في مقالة هذا المجلد حالتين ينشأ عنهما ارتباط وعدم استقلال مجتمعين للدراسة. في الحالة الأولى يكون المجتمع الأول هو القوم التي يمكن مشاهدتها على المفردات قبل تطبيق معالجة معينة، بينما يكون المجتمع الثاني هو القيم التي يمكن مشاهدتها على نفس المفردات بعد تطبيق المعالجة. وفي الحالة الثانية، يتم تقسيم مفردات مجتمع الدراسة إلى أزواج تتماثل وتتشابه في خصائصها حيث تخضع إحدى المفردتين في كل زوج عشوائياً لمعاملة معينة بينما تخضع المفردة الثانية لمعاملة أخرى.

كما في حالة الاستقلال يكون مقدار نقطة لمعلمة الفرق بين متوسطي المجتمعين، $\Delta = \mu_1 - \mu_2$ ، هو الفرق بين متوسطي العينتين، أو متوسط الفرق بين المشاهدات المتناظرة في العينتين. وإذا رمزنا

لملاحظات العينة الأولى بالرمز s ولملاحظات العينة الثانية بالرمز s' ،
تحتسب فروق الملاحظات ونرمز لها بالرمز f .

س	ص	ف = س - ص
س _١	ص _١	ف _١ = س _١ - ص _١
س _٢	ص _٢	ف _٢ = س _٢ - ص _٢
...
...
س _ن	ص _ن	ف _ن = س _ن - ص _ن

ويتم إجراء اختبارات الفروض وتقدير فترات الثقة باستخدام
الفروق السابقة، ونقوم بتطبيق أساليب الاستدلال المتعلقة بمتوسط مجتمع
واحد (مجتمع الفروق بين مشاهدات المجموعتين) والتي سبق أن تناولناها
في الفصلين الثاني والثالث.

مقدر النقطة

$$\hat{\Delta} = \bar{f} = \frac{\text{مجموع } f}{n} = \bar{s} - \bar{v}$$

وإذا كان لكل من مجتمعي الدراسة توزيع طبيعي، يكون لمجتمع الفروق
توزيع طبيعي أيضا متوسطه هو Δ ونرمز لتباينه بالرمز σ_f^2 . وهناك
علاقة بين تباين مجتمع الفروق، σ_f^2 ، وتبايني المجتمعين σ_s^2 و σ_v^2
بالإضافة إلى معمة أخرى يطلق عليها اسم التباين بين متغيري الدراسة،

ولكننا لن نستخدم هذه العلاقة لإيجاد مقدر نقطة لمعلمة تباين الفرق نظرا لعدم دراسة التغيرات بعد، ولكننا سوف نقوم هنا بتقدير تباين الفرق باستخدام تباين الفروق بين مشاهدات العينتين.

$$\hat{\sigma}_f^2 = \hat{\sigma}_E^2 = \frac{1}{n-1} \text{مجم} (f - \bar{f})^2$$

أو

$$\hat{\sigma}_f^2 = \frac{1}{n-1} \left[\text{مجم} f^2 - \frac{(\text{مجم} f)^2}{n} \right]$$

وعندما يكون حجم العينتين (ودائما ما يكون متساوي نظرا لارتباط المشاهدات في هذه الحالة) صغيرا، فإننا نستخدم توزيع t لتقدير فترات الثقة واختبارات الفروض كما سبق. ويكون مقدر فترة ثقة $1-\alpha$ % للفرق بين متوسطي المجتمعين الأول والثاني

$$\bar{f} \pm t_{\alpha/2, n-1} \times \frac{\hat{\sigma}_f}{\sqrt{n}}$$

وعند إجراء اختبارات الفروض، إذا كتبنا الفروض الإحصائية بدلالة المعلمة Δ تكون الفروض على إحدى الصور

اختبار الطرفين:

$$H_0: \Delta = \Delta_0 \quad \text{مقابل} \quad H_1: \Delta \neq \Delta_0$$

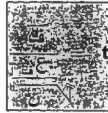
اختبار الطرف الأيمن:

$$H_0: \Delta \leq \Delta_0 \quad H_1: \Delta > \Delta_0$$

اختبار الطرف الأيسر

$$H_0: \Delta \geq \Delta_0 \quad H_1: \Delta < \Delta_0$$

إحصائية الاختبار



القيمة الحرجة

يتم تحديد القيمة الحرجة، في حالة العينات صغيرة الحجم، من جدول توزيع t عند درجات حرية $(n - 1)$ والاحتمال المناسب حسب نوع الاختبار. وفي حالة العينات كبيرة الحجم نستخدم جداول التوزيع الطبيعي المعياري.

نتيجة الاختبار

إذا زادت القيمة العددية لإحصائية الاختبار عن القيمة الجدولية، نقوم برفض فرض العدم وقبول الفرض البديل ولستحتاج وجود فروق جوهرية بين بيانات مجموعتي الدراسة.

مثال ٤-٦

لقياس مدى فاعلية برنامج تدريبي معين لرفع الكفاءة الإنتاجية للعاملين، جمعت البيانات التالية عن زمن إنتاج (بالمساعات) الوحدة الواحدة من منتج معين لكل عامل في عينة عشوائية قبل وبعد تنفيذ البرنامج عليهم.

رقم العامل	١	٢	٣	٤	٥	٦	٧	٨
الزمن قبل	٦	٥,٥	٥	٥	٦	٥,٨	٦,٥	٦
الزمن بعد	٥,٤	٥	٤,٦	٥	٥,٥	٥,٥	٥,٨	٦,٢

وعلى فرض أن توزيع زمن الانتاج قبل وبعد تدريب العمال هو توزيع طبيعي، فالمطلوب:

1- إيجاد تقدير فترة ثقة 95% للفرق بين متوسطي زمن إنتاج الوحدة

قبل البرنامج وبعده إذا ما طبق على جميع العاملين.

2- عند مستوى المعنوية 5%، هل توصي بتطبيق هذا البرنامج

التدريبي على كافة العاملين؟

الحل:

افترض أن مجتمع أزمنة الانتاج قبل البرنامج هو المجتمع الأول ومتوسطه

μ_1 ، وأن مجتمع أزمنة الانتاج بعد البرنامج هو المجتمع الثاني ومتوسطه

μ_2 وأن الفرق بين المتوسطين هو :

$$\mu_2 - \mu_1 = \Delta$$

ويجب مراعاة الاتجاه السابق عند إيجاد الفروق بين مشاهدات العينتان حيث نطرح، حسب تعريف Δ ، قيم زمن الانتاج بعد البرنامج من القيم المناظرة قبل البرنامج، ثم نوجد الوسط الحسابي والانحراف المعياري للفروق.

٦	٦,٥	٥,٨	٦	٥	٥	٥,٥	٦	الزمن قبل
٦,٢	٥,٨	٥,٥	٥,٥	٥	٤,٦	٥	٥,٤	الزمن بعد
٠,٢-	٠,٧	٠,٣	٠,٥	٠	٠,٤	٠,٥	٠,٦	ف
٠,٠٤	٠,٤٩	٠,٠٩	٠,٢٥	٠	٠,١٦	٠,٢٥	٠,٣٦	ف ^٢

$$\bar{f} = \frac{2,8}{8} = 0,35$$

$$E_f^2 = \frac{1}{1-n} \left[\text{مجم ف}^2 - \frac{(\text{مجم ف})^2}{n} \right]$$

$$= \frac{1}{7} \left[\frac{(2,8)^2}{8} - 1,64 \right]$$

$$= 0,0943 = 0,66 \times \frac{1}{7}$$

$$E_f = 0,307$$

1- تقدير فترة ثقة 95% للفروق بين متوسطي الزمن قبل وبعد

تطبيق البرنامج التدريبي

$$0.025 = 2\alpha \quad 0.05 = \alpha \quad 0.95 = 1 - \alpha$$

$$2,360 = (0.025, \eta) t \quad Y = 1 - 8 = 3$$

$$\frac{\bar{C}}{\sqrt{n}} \times (2\alpha, 1-\alpha) t \pm \bar{C}$$

$$\frac{0,307}{8\sqrt{}} \times 2,360 \pm 0,30$$

$$0,207 \pm 0,30$$

$$0,093 = 0,207 - 0,30 = \text{الحد الأدنى}$$

$$0,607 = 0,207 + 0,30 = \text{الحد الأعلى}$$

2- إعطاء توصية بتطبيق البرنامج أو بعدم تطبيقه، نقوم بإجراء اختبار للفروض يوضح ما إذا كان البرنامج التدريبي ذو جدوى وفائدة أم لا. وفي ضوء المعلومات المتاحة، يكون البرنامج فعال إذا كان يؤدي إلى تخفيض زمن الانتاج بصورة حقيقية وجوهرية. أي أن البرنامج يكون فعال إذا كانت القيمة الحقيقية لمعلمة الفرق Δ ، وفقا لتعريفنا لها، موجبة (الفرض البديل). وعلى الجانب الآخر لا يكون البرنامج فعالا إذا بقي متوسط زمن الانتاج كما هو وأسوأ من ذلك إذا زاد زمن الانتاج بعد التدريب، وهي الحالة التي تكون فيها قيمة Δ أقل من أو تساوي الصفر (فرض العدم). وعلى ذلك يكون الاختبار في هذه الحالة هو اختبار طرف ليمين.

الفروض الإحصائية

إحصائية الاختبار $H_0: \Delta \geq \text{صفر}$ مقابل $H_1: \Delta < \text{صفر}$.

$$\frac{\bar{f} - \bar{f}_c}{\frac{s_c}{\sqrt{n}}} = t^*$$

$$2,220 = \frac{0,30}{\frac{0,307}{\sqrt{8}}} =$$

القيمة الحرجة

$$1,890 = (0.05, 7)t \quad 0.05 = \alpha \quad v = 7$$

نتيجة الاختبار

حيث أن قيمة إحصائية الاختبار تزيد عن القيمة الحرجة، فإننا نرفض فرض العدم ونقبل الفرض البديل ونستنتج أن متوسط زمن الانتاج بعد برنامج التدريب يقل بصورة جوهرية عما كان عليه قبل البرنامج. بناءً على ذلك يجب التوصية بتطبيق هذا البرنامج على جميع العاملين وذلك عند مستوى معنوية ٥%.

٤-٤ أساليب الاستدلال الإحصائي عن الفرق

يبين نسبتين في مجتمعين مستقلين

درسنا في الفصلين السابقين أساليب تقدير فترات الثقة واختبارات الفروض الإحصائية المتعلقة بنسبة ما في مجتمع إحصائي معين، وقد وجدنا أننا نعتمد على نظرية النهاية المركزية للحصول على توزيع النسبة في المجتمع. وتظهر الحاجة في العديد من الدراسات إلى إجراء اختبارات للفروض تتعلق بمعنوية الفرق بين نسبتين في مجتمعين مستقلين وإلى تقدير حدود ثقة للفرق بين النسبتين. ونتناول هذه المشكلة من مشاكل الاستدلال الإحصائي في هذا المبحث .

يتم الاستدلال عن الفرق بين نسبتين في مجتمعين مستقلين باستخدام عينتين مسحوبتين منهما. وإذا كان حجم العينة الأولى هو n_1 مفردة من بينها k_1 مفردة تحقق خاصية الدراسة، وكان حجم العينة الثانية هو n_2 ومن بينها k_2 مفردة تحقق خاصية الدراسة، فإننا نعرف بعض الرموز والمقاييس المستخدمة في عملية الاستدلال على الوجه التالي:

المجتمع الأول	المجتمع الثاني	النسبة في المجتمع
1θ	2θ	

النسبة في العينة	$ق_1 = \frac{1ك}{1ن}$	$ق_2 = \frac{2ك}{2ن}$
------------------	-----------------------	-----------------------

النسبة المشتركة	$ق_م = \frac{1ك + 2ك}{1ن + 2ن}$
	$= \frac{1ق_1 \times 1ن + 2ق_2 \times 2ن}{1ن + 2ن}$

٤-٤-١ مقدار نقطة للفرق بين نسبتي المجتمعين

نستخدم الفرق بين نسبتي العينتين كمقدّر نقطة للفرق بين نسبتي المجتمعين. فإذا حسبنا الفرق بين نسبتي المجتمعين بطرح نسبة المجتمع الثاني من نسبة المجتمع الأول، $1\theta - 2\theta$ ، يكون مقدار نقطة لمعلمة الفرق هو $ق_1 - ق_2$.

٤-٤-٢ توزيع المعاينة للفرق بين نسبتي عينتين مستقلتين

من نظرية النهاية المركزية في الفصل الأول، نعلم أن التوزيع

التقريبي للنسبة $ق_1$ يكون توزيع طبيعي ع(1θ ، $\frac{(1\theta - 1)1\theta}{1ن}$) ويكون

التوزيع التقريبي للنسبة $ق_2$ هو توزيع طبيعي ع(2θ ، $\frac{(2\theta - 1)2\theta}{2ن}$)

وذلك عندما يكون حجم كل عينة من العينتين كبيراً بدرجة كافية. ويمكننا من هذه المعلومات استخدام نظرية ١-٤ من الفصل الأول لنجد أن التوزيع التقريبي للفرق بين نسبتي العينتين هو توزيع طبيعي:

$$E \left(\frac{(\bar{p}_1 - \bar{p}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}} \right) \approx 0$$

وبالتالي يكون للمتغير العشوائي:

$$Z = \frac{(\bar{p}_1 - \bar{p}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}} \approx 0$$

توزيع طبيعي معياري ع (صفر ، 1) تقريباً.

٤-٤-٣ تقدير فترات الثقة للفرق بين نسبتي

كما نعلم يمثل المقام في الصيغة السابقة الخطأ المعياري لتوزيع المعاينة للفرق بين نسبتي العينتين. وبضربه في قيمة $Z_{1-\alpha/2}$ نحصل على حد خطأ التقدير. بالتالي نحصل مقدار فترة ثقة $(1 - \alpha)\%$ للفرق بين نسبتي المجتمعين $(\theta_1 - \theta_2)$ بإضافة وطرح حد خطأ التقدير عند احتمال $(1 - \alpha)$ إلى مقدار النقطة للفرق بين النسبتين.

$$\frac{(1-\theta_1)\theta_1}{n_1} + \frac{(1-\theta_2)\theta_2}{n_2} \sqrt{r_{1-\alpha} Z \pm (q_2 - q_1)}$$

ويلاحظ أننا لا نستطيع حساب حدي فترة الثقة من الصيغة السابقة لتضمنها المعلمتين θ_1 و θ_2 المجهولتين، وبالتالي فإننا نستخدم مقدر النقطة لكل منهما لنحصل على صيغة مقدر فترة الثقة على الصورة:

$$\frac{(1-\hat{\theta}_1)\hat{\theta}_1}{n_1} + \frac{(1-\hat{\theta}_2)\hat{\theta}_2}{n_2} \sqrt{r_{1-\alpha} Z \pm (q_2 - q_1)}$$

٤-٤-٤ اختبارات الفروض لمعنوية الفرق بين نسبتي

الفروض الإحصائية

إذا أردنا اختبار معنوية الفرق بين نسبتي في مجتمعين وما إذا كانت إحدى النسبتين تزيد عن الأخرى أم لا، فإن الفروض الإحصائية تكون على أحد الصور الثلاث كما سبق.

اختبار الطرفين

$$H_0: \theta_1 = \theta_2 \quad \text{مقابل} \quad H_1: \theta_1 \neq \theta_2$$

أو

$$H_0: \theta_1 = \theta_2 \quad \text{ضد} \quad H_1: \theta_1 < \theta_2 \quad \text{ضد} \quad H_1: \theta_1 > \theta_2$$

اختبار الطرف الأيمن

$$\theta \geq \theta_0 : H_0 \quad \text{مقابل} \quad \theta < \theta_0 : H_1$$

أو

$$\theta \geq \theta_0 : H_0 \quad \text{صفر} \quad \text{مقابل} \quad \theta < \theta_0 : H_1 \quad \text{صفر}$$

اختبار الطرف الأيسر

$$\theta \leq \theta_0 : H_0 \quad \text{مقابل} \quad \theta > \theta_0 : H_1$$

أو

$$\theta \leq \theta_0 : H_0 \quad \text{صفر} \quad \text{مقابل} \quad \theta > \theta_0 : H_1 \quad \text{صفر}$$

إحصائية الاختبار

حيث أننا نجري اختبارات الفروض دائما على أساس صحة فرض العدم، والذي يتضمن تساوي النسبتين في المجتمعين، أو أن الفرق بينهما يساوي الصفر، فإننا نقوم بوضع $(\theta_0 - \theta) = \text{صفر}$ في صيغة Z^* السابقة. من جهة أخرى، نقوم في المقام باستخدام النسبة المشتركة لتقدير النسبتين في صيغة الخطأ المعياري على أساس تساويهما في ظل صحة فرض العدم. من هذا تكتب صيغة إحصائية الاختبار على الصورة:

$$Z = \frac{\bar{y} - \theta_0}{\sqrt{\frac{1}{n} \left(\frac{1}{\theta_0} + \frac{1}{\theta_1} \right)}}$$

القيمة الحرجة

تحدد القيمة الحرجة على أساس نوع الاختبار من جداول التوزيع الطبيعي المعياري، وهي كما سبق $Z_{1-\alpha/2}$ في حالة اختبار الطرفين ، $Z_{1-\alpha}$ في حالة اختبار الطرف الأيمن و $-Z_{1-\alpha}$ في حالة اختبار الطرف الأيسر.

نتيجة الاختبار

إذا وقعت قيمة إحصائية الاختبار داخل المنطقة الحرجة، نرفض فرض العلم ونقبل الفرض البديل ونستنتج وجود اختلاف بين النسبتين في المجتمعين، أو أن أحدهما تزيد على الأخرى.

مثال ٤-٧

تم سؤال عينة عشوائية من ٢٠٠ طالب وعينة عشوائية من ١٥٠ طالبة عن مدى التزامهم بحضور جميع المحاضرات أثناء تواجدهم داخل الحرم الجامعي، فوجد أن ١٢٤ طالبا و ١٠٥ طالبات من طلبة العينة يلتزمون بحضور جميع المحاضرات.

1- أوجد تقدير فترة ثقة 95% للفرق بين نسبتي الطلاب والطالبات الملتزمون بحضور المحاضرات.

2- عند مستوى معنوية 5%، هل تعطي هذه المعلومات دلالة كافية على أن الطالبات أكثر التزاما من الطلاب بحضور المحاضرات؟

الحل:

طلاب	طلّبات	
n_2	n_1	النسبة في المجتمع
$200 = n_2$	$100 = n_1$	حجم العينة
$124 = x_2$	$100 = x_1$	عدد الطلبة الموافقين

$$\text{النسبة في العينة} \quad q_1 = \frac{100}{100} = 1, \quad q_2 = \frac{124}{200} = 0,62$$

$$\text{النسبة المشتركة} \quad q_m = \frac{124 + 100}{200 + 100} = 0,604$$

تقدير فترة ثقة 95% للفرق بين النسبتين

$$1.96 = 0.975Z = 2\alpha - 1Z$$

$$(q_1 - q_2) \pm Z_{1-\alpha/2} \sqrt{\frac{q_1(1-q_1)}{n_1} + \frac{q_2(1-q_2)}{n_2}}$$

$$(0,62 - 1,00) \pm 1,96 \sqrt{\frac{0,62 \times 0,38}{200} + \frac{1,00 \times 0,00}{100}}$$

$$\pm 0,0995$$

$$-0,0995 = -0,10 \quad \text{الحد الأدنى}$$

$$+0,0995 = 0,10 \quad \text{الحد الأعلى}$$

2- هل الطالبات أكثر مواظبة على حضور المحاضرات؟

للإجابة على هذا السؤال نجري اختباراً للفروض يتضمن فيه فرض العدم أن الطالبات لسن أكثر التزاماً وأكثر مواظبة على حضور المحاضرات (نسبة الإثبات المواظبات على الحضور لا تختلف جوهرياً عن نسبة الذكور، وقد تكون أقل). من جهة أخرى نضع الفرض البديل على أساس أن نسبة الإثبات تزيد على نسبة الذكور. بالتالي وعلى ضوء تعريفنا للمجتمع الأول والمجتمع الثاني يكون الاختبار ذو طرف أيمن وتكون الفروض الإحصائية على الصورة:

$$H_0: \theta \geq \theta_0 \quad \text{مقابل} \quad H_1: \theta < \theta_0$$

$$\text{أو} \quad H_0: \theta - \theta_0 \geq \text{صفر} \quad \text{مقابل} \quad H_1: \theta - \theta_0 < \text{صفر}$$

إحصائية الاختبار:

$$\begin{aligned} & \frac{0,62 - 0,70}{\left(\frac{1}{200} + \frac{1}{150}\right)(0,346)} = \\ & 0,05 = \frac{0,08}{0,0014} = \end{aligned}$$

القيمة الحرجة

اختبار طرف واحد عند مستوى معنوية $\alpha = 0.05$ من جدول النسب المختارة للتوزيع الطبيعي المعياري $Z_{\alpha-1} = Z_{0.95} = 1.645$

نتيجة الاختبار

بمقارنة قيمة إحصائية الاختبار بالقيمة الحرجة، نجد أن قيمة إحصائية الاختبار تقل عن القيمة الحرجة، $1,00 > 1,640$. بالتالي لا يمكننا رفض فرض العدم والذي يتضمن عدم وجود اختلاف جوهري بين النسبتين في المجموعتين . وذلك عند مستوى معنوية $0,05$

٤-٥ اختبارات الفروض بتساوي تبايني مجتمعين مستقلين إلهما توزيع طبيعي

رأينا في المبحث الثاني من هذا الفصل أن إحصائية اختبار الفرض بتساوي متوسطي مجتمعين مستقلين تعتمد على ما إذا كان تبايني المجتمعين متساويين أم لا، حيث نقوم في حالة تساوي التباينين بحساب تباين مشترك واحد كمتوسط مرجح لتبايني العينتين. ومن هذا تظهر أهمية دراسة اختبار للفروض يتعلق بتساوي تبايني مجتمعين طبيعيين ومستقلين، وهو ما نقوم به في هذا المبحث. ويشتمل أسلوب الاختبار على المراحل الأربع المعتادة لاختبارات الفروض.

الفروض الإحصائية

إذا رمزنا لتباين المجتمع الأول بالرمز σ_1^2 ولتباين العينة المسحوبة منه بالرمز σ_1^2 ، وإذا رمزنا لتباين المجتمع الثاني بالرمز σ_2^2

ولتباين العينة المسحوبة منه بالرمز σ_1^2 ، تكون الفروض الإحصائية على إحدى الصورتين:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad \text{مقابل} \quad H_1: \sigma_1^2 < \sigma_2^2$$

أو

$$H_0: \sigma_1^2 \geq \sigma_2^2 \quad \text{مقابل} \quad H_1: \sigma_1^2 < \sigma_2^2$$

ويمكن إعادة كتابة الفروض بصورة بديلة على أساس النسبة بين التباينين، وليس الفرق بينهما كما في حالتنا اختبارات الفروض للفرق بين وسطين حسابيين أو الفرق بين نسبتيين.

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \quad \text{مقابل} \quad H_1: \frac{\sigma_1^2}{\sigma_2^2} < 1$$

أو

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} \geq 1 \quad \text{مقابل} \quad H_1: \frac{\sigma_1^2}{\sigma_2^2} < 1$$

يلاحظ من هذا أننا سوف نجري اختبار تساوي تبايني مجتمعين على صورة اختبار طرف أيمن. ويتم اختيار أحد الصورتين البديلتين للفروض بالنظر إلى تبايني العينتين بحيث يظهر التباين الأكبر دائما في البسط. فإذا وجدنا

أن \bar{X}_1 أكبر من \bar{X}_2 ، كتبنا الفرض البديل على الصورة $\frac{\bar{X}_1}{\bar{X}_2} < 1$. وإذا

وجدنا أن \bar{X}_2 أكبر من \bar{X}_1 ، كتبنا الفرض البديل على الصورة $\frac{\bar{X}_1}{\bar{X}_2} < 1$.

إحصائية الاختبار

نقوم هنا باستخدام تباين كل عينة كمقدر نقطة لتباين المجتمع المسحوبة منه وتكون إحصائية الاختبار هي النسبة بين تبايني العينتين بحيث تكون هذه النسبة أكبر من الواحد الصحيح. وسوف نرمز لإحصائية الاختبار بالرمز F^* .

$$F^* = \frac{\bar{X}_1}{\bar{X}_2} \quad \text{إذا كانت } \bar{X}_1 \text{ أكبر من } \bar{X}_2$$

$$F^* = \frac{\bar{X}_2}{\bar{X}_1} \quad \text{إذا كانت } \bar{X}_2 \text{ أكبر من } \bar{X}_1$$

وعندما يكون تباين كل عينة مقدر جيد لتباين المجتمع، أي يعبر عنه بصورة صادقة، فإن قيمة إحصائية الاختبار سوف تكون قريبة من الواحد الصحيح عندما يكون فرض العدم صحيحا. ومن ناحية أخرى في حالة عدم صحة فرض العدم، نتوقع أن يكون تباين العينة الذي يظهر في البسط أكبر بصورة جوهرية من تباين العينة الذي يظهر في المقام. ويتضمن هذا أن

تكون قيمة إحصائية الاختبار أكبر كثيرا من الواحد الصحيح.

القيمة الحرجة

لتحديد القيمة الحرجة التي تفصل بين منطقتي قبول ورفض فرض العدم، يلزم تحديد توزيع المعاينة لإحصائية الاختبار والتي تمثلها النسبة بين تبايني العينتين. ودون الخوض في مزيد من التفاصيل، التي تفوق مستوى الدارسين لهذا الكتاب، نقول بأن توزيع النسبة بين تباينين مستقلين محسوبين من بيانات عينات مسحوبة من مجتمعات طبيعية، وفي ظل صحة فرض العدم بتساوي التباينين، يكون توزيعا احتماليا يطلق عليه اسم توزيع النسبة F . ويكون لتوزيع F عدنان لدرجات الحرية، درجات حرية البسط (د) ودرجات حرية المقام (ر). وتوجد جداول لتوزيعات F كل منها يخص أحد قيم مستوى المعنوية شائعة الاستخدام. يوجد بنهاية هذا الكتاب جدول لمستوى المعنوية $\alpha = 0.05$ ، ويوجد جنول آخر لمستوى المعنوية $\alpha = 0.01$. ويشتمل الصف العلوي في الجدول على درجات حرية البسط ، بينما يشتمل العمود الأول على درجات حرية المقام. وعادة ما نكتب القيم المستخرجة من جداول توزيع F على الصورة $F_{(د، ر، \alpha)}$. على سبيل المثال، يتم استخراج قيمة $F_{(8، 10، 0.05)}$ من الجدول الخاص بمستوى المعنوية 5% بالكشف أسفل درجات حرية البسط (8) وأمام درجات حرية المقام (10). ومن الجدول نجد أن:

$$F_{(8، 10، 0.05)} = 3.07$$

كذلك يتم استخراج قيمة $F(12, 6, 0.01)$ من الجدول الخاص بمستوى المعنوية 1% بالكشف أسفل درجات حرية البسط (12) وأمام درجات حرية المقام (6). ومن الجدول نجد أن:

$$F(12, 6, 0.01) = 7.72.$$

وفي اختبارات الفروض المتعلقة بتساوي تبايني مجتمعين، إذا كانت

$$F = \frac{\frac{1}{2} \frac{E}{\chi^2}}{\frac{1}{2} \frac{E}{\chi^2}} \text{ تكون القيمة الحرجة هي}$$

$$F(\alpha, 1 - \alpha, 1 - \alpha, 1 - \alpha, n) \cdot \text{وإذا كانت إحصائية الاختبار على الصورة}$$

$$F = \frac{\frac{1}{2} \frac{E}{\chi^2}}{\frac{1}{2} \frac{E}{\chi^2}} \text{ تكون القيمة الحرجة هي } F(\alpha, 1 - \alpha, 1 - \alpha, 1 - \alpha, n)$$

نتيجة الاختبار

إذا زالت قيمة إحصائية الاختبار عن القيمة الحرجة، نرفض فرض العدم ونقبل الفرض البديل ونستنتج أن المجتمع الذي ظهر تباين عينته في بسط إحصائية الاختبار يزيد بصورة جوهرية عن تباين المجتمع الآخر. وفي هذه الحالة عند إجراء اختبار تساوي متوسطي المجتمعين لا نقوم بحساب التباين المشترك. على الجانب الآخر، إذا قلت قيمة إحصائية الاختبار عن القيمة الحرجة، فإن هذا يعد دلالة على تساوي تبايني المجتمعين، وبالتالي

نقوم باستخدام التباين المشترك عند إجراء اختبار للفروض بتساوي متوسطي المجتمعين.

مثال ٤-٨

افترضنا في مثال ٤-٤ أن تبايني مجتمعي الذكور والإناث متساويين. والمطلوب اختبار صحة هذا الافتراض عند مستوى معنوية 1%.

الحل:

من بيانات مثال ٤-٤ نجد أن

إناث	ذكور	
$\frac{1}{2}\sigma$	$\frac{1}{2}\sigma$	تباين المجتمع
$\frac{1}{2}\sigma^2 = 0,64$	$\frac{1}{2}\sigma^2 = 1$	تباين العينة
$n = 2$	$n = 10$	حجم العينة

حيث أن تباين عينة الذكور أكبر من تباين عينة الإناث، تكون الفروض الإحصائية على الصورة:

$$H_0: \frac{\frac{1}{2}\sigma^2}{\frac{1}{2}\sigma^2} = 1 \quad \text{مقابل} \quad H_1: \frac{\frac{1}{2}\sigma^2}{\frac{1}{2}\sigma^2} < 1$$

إحصائية الاختبار

$$\frac{24}{24} = F^*$$

$$\frac{1}{0.64} =$$

$$1.5625 =$$

يتم استخراج القيمة الحرجة من جدول توزيع F الخاص بمستوى المعنوية 1% بالكشف أسفل درجات حرية البسط وهي حجم العينة الأولى مطروحا منه واحد (ن₁ - 1 = 9) (لأنها ظهرت في بسط الإحصائية) وأمام درجات حرية المقام وهي حجم العينة الثانية مطروحا منه واحد (ن₂ - 1 = 7).

$$F(9, 7, 0.01) = 6.72.$$

نتيجة الاختبار

حيث أن قيمة إحصائية الاختبار تقل كثيرا عن القيمة الحرجة، فإننا نقبل فرض العدم ونستنتج أنه يمكن أن يكون لمجتمع توزيع درجات الذكور وتوزيع درجات الإناث تباين متساوي، أو لا يوجد فرق جوهري بين التباينين. لذلك عند إجراء اختبار للفروض يتعلق بمتوسطي المجتمعين، قمنا باستخدام التباين المشترك.

تمارين الفصل الرابع

(1) سحبت عينة عشوائية حجمها ١٢ مفردة من مجتمع له توزيع طبيعي ع(١١، ٢٤) ، كما سحبت عينة عشوائية حجمها ٢٠ مفردة من مجتمع آخر، مستقل عن المجتمع الأول وله توزيع طبيعي ع(١١، ١٥) .

1- اكتب مقدر نقطة للفرق بين متوسطي المجتمعين.

2- ما هو توزيع المعاينة لمقدر النقطة في المطلوب السابق؟

3- ما هو احتمال أن يكون خطأ تقدير مقدر النقطة للفرق بين متوسطي المجتمعين هو ثلاث وحدات على الأكثر.

4- على فرض أن متوسط المجتمع الأول يزيد عن متوسط المجتمع الثاني بثلاث وحدات، ما هو $P(|\bar{X}_1 - \bar{X}_2| \geq 2)$ ؟

5- نظرا لكبر حجم المجتمع الثاني وكبر أهميته النسبية نريد اختيار عينتين أخريتين من المجتمعين بحيث يكون حجم عينة المجتمع الأول نصف حجم عينة المجتمع الثاني، ما هو حجم كل من العينتين الذي يجعل أقصى خطأ لتقدير الفرق بين

متوسطي المجتمعين باستخدام الفرق بين متوسطي العينتين
وحدة واحدة باحتمال 95%.

(2) لقياس تأثير الموقع على حجم مبيعات المحلات التجارية، تم اختيار
عينة عشوائية من ١٠ محلات للملابس الجاهزة تقع جميعها داخل
مراكز للتسوق وعينة أخرى من ١٥ محلا منتشرة في الأحياء السكنية
وتم تسجيل مبيعات كل محل منها على مدار أسبوع، فوجد أن متوسط
حجم المبيعات اليومية خلال فترة الدراسة لمحلات مراكز التسوق هو
٣٤٥٠ جنيهها باحتراف معياري ٨٠ جنيهها. كذلك وجد أن متوسط
المبيعات اليومية للمجموعة الأخرى هو ٣١٢٥ جنيهها باحتراف معياري
٦٠ جنيهها. على فرض أن قيم المبيعات اليومية لمحلات الملابس
الجاهزة تتبع توزيع طبيعي تقريبا ونفس التباين فالمطلوب:

1- عند مستوى معنوية 5%، اختبر تأثير الموقع على حجم
المبيعات.

2- اكتب تقدير فترة ثقة 95% للفرق بين متوسطي قيم المبيعات
اليومية لمحلات الملابس الجاهزة في كلا الموقعين.

(3) تم تدريب عينة من ٢٠ رجلا و٢٥ سيدة على إنجاز مهمة معينة، ثم
طلب من كل منهم القيام بإتجاز تلك المهمة وتم قياس الزمن فكان
متوسط الزمن للرجال هو ٥ دقائق والسيدات ٤ دقائق. وإذا علمت أن

توزيع زمن إنجاز هذه المهمة يتبع بصفة عامة توزيع طبيعي تقريبا
بالتحرف معياري متساوي قدره ٧٢ ثانية لكلا الرجال والسيدات.

1- عند مستوى معنوية 1% هل تؤيد هذه البيانات أن السيدات
أكثر كفاءة من الرجال في إنجاز المهمة؟

2- أوجد تقدير فترة ثقة 98% للفرق بين متوسطي زمن الرجال
والسيدات.

(4) لاختبار تأثير نوع جديد من أدوية علاج مرض ارتفاع ضغط الدم، تم
اختيار عينة من ٢٠ مريضاً في نفس الحالة الصحية ونفس السن
تقريباً وتم تقسيمهم إلى مجموعتين إحداهما تتكون من ١٢ مريضاً تم
إعطائهم أقراصاً خالية من المادة الفعالة بينما تم إعطاء الثمانية
المتبقين الدواء الجديد. وبعد فترة من تناول العلاج تم قياس ضغط الدم
لكل من المرضى فكانت بيانات بسط الضغط كما يلي:

المجموعة الأولى: ١٨٠، ١٩٠، ١٧٥، ١٩٠، ٢١٠، ١٨٤، ١٧٣،
١٨٠، ١٨٣، ٢٠٠، ١٩٥، ١٦٥.

المجموعة الثانية: ١٤٠، ١٦٥، ١٥٠، ١٥٥، ١٤٠، ١٨٠، ١٧٠، ١٥٠.

وعلى فرض أن توزيع ضغط الدم لدى المرضى هو توزيع طبيعي بنفس
التباين للمجموعتين

(أ) علق على فاعلية الدواء الجديد عند مستوى معنوية 1%.

(ب) اكتب تقدير فترة ثقة 99% للفرق بين متوسطي مقياس الضغط للمرضى الذين لا يتناولون أي علاج والمرضى الذين يتناولون الدواء الجديد.

(5) لقياس مدى فاعلية نظام جديد لإنقاص الوزن في أحد المعاهد الرياضية، تم تحديد أوزان ثمانية أشخاص قبل تطبيق هذا النظام وتحديد أوزانهم بعد مضي شهرين على تطبيق هذا النظام فكانت البيانات كما يلي:

الوزن قبل النظام: 80، 85، 78، 84، 87، 96، 92، 84

الوزن بعد النظام: 77، 84، 76، 81، 82، 92، 90، 80

(أ) اختبر معنوية تأثير النظام على إنقاص الوزن عند مستوى معنوية 5%.

(ب) أوجد تقدير فترة ثقة 95% للفرق بين متوسطي الوزن قبل وبعد تطبيق للنظام.

(6) اتتجت إحدى شركات البترول مادة جديدة وادعت أنها إذا ما أضيفت إلى وقود السيارة فأنها تزيد من عدد الكيلومترات التي تقطعها السيارة لكل لتر مستهلك من الوقود. ولتجربة تأثير هذه المادة تم اختيار عينة عشوائية من ست سيارات وتم تسيرها لمدة أسبوع باستخدام وقود عادي وبدون إضافة المادة، ثم تم تسير نفس السيارات لمدة أسبوع

آخر مع إضافة المادة لنفس نوع الوقود. وقد سجلت البيانات التالية
لمتوسط عدد الكيلومترات التي قطعتها كل سيارة لكل لتر خلال
الأسبوعين:

بدون المادة: ٥,٧٥ ، ٦,٢٥ ، ٥ ، ٦ ، ٥,٥ ، ٧

بإضافة المادة: ٦,٢٥ ، ٦ ، ٥,٥ ، ٧ ، ٦,٥ ، ٨,٢٥

(أ) هل تنصح قائدي السيارات بإضافة المادة الجديدة عند مستوى
معنوية 5%؟

(ب) أوجد تقدير فترة ثقة 95% للفرق بين متوسطي عدد الكيلومترات
لكل لتر عند إضافة المادة وعند عدم إضافتها.

(7) لدى إحدى الشركات سلسلة مطاعم في مدينتي (أ ، ب) وفي عينة من
٢٠٠ عميل في المدينة (أ) وجد أن هناك ١٢٠ شخصاً يعتقدون أن
مستوى الخدمة ممتاز. بينما في عينة من ٢٥٠ عميل في المدينة (ب)
وجد أن عدد العملاء الذين يعطون نفس التقدير للخدمة هو ١٢٥.

أولاً: هل تعطي هذه المعلومات دلالة على عدم وجود اختلاف بين
مستوى خدمة المطاعم في المدينتين من وجهة نظر العملاء عند
مستوى معنوية 5%؟

ثانياً: اكتب تقدير فترة ثقة 95% للفرق بين نسبتي العملاء الذين يعتقدون أن مستوى الخدمة ممتاز في المدينتين.

(8) في عينة من 400 طالب و 200 طالبة في الجامعة وجد أن عدد من استخدم شبكة المعلومات The Internet مرة واحدة على الأقل هو 112 طالباً و 46 طالبة.

اختبر الفرض بأن نسبة مستخدمي الشبكة من الطلاب أعلى من نسبة الطالبات عند مستوى معنوية 5% ، ثم أوجد تقدير فترة ثقة 95% للفرق بين النسبتين.

الفصل الخامس

تحليل التباين

Analysis of Variance

(٥ - ١) مقدمة :

سنعرض في هذا الفصل لكيفية المقارنة بين أكثر من متوسط مجتمعين باستخدام البيانات التي تُجمع وفقاً لتصميم تجريبي معين .

وقد كان للتجارب الزراعية دوراً رائداً في مجال تصميم التجارب Experimental design لذلك نجد أن مصطلحات هذا الموضوع يغلب عليها الطابع الزراعي . فمعظم التجارب الزراعية تتضمن معالجة الوحدات التجريبية بطريقتين أو أكثر ثم المقارنة بين متوسطات المجتمعات المناظرة للمعالجات المختلفة . على سبيل المثال المقارنة بين متوسطات إنتاجية القدان في مجموعة من حقول القمح تم زراعتها باستخدام أربعة أنواع مختلفة من الأسمدة . المقارنة بين متوسطات الزيادة في أوزان مجموعة من الأبقار الناتجة عن اتباع أربعة أنواع من أنظمة التغذية للتسمين . ويطلق على كل من أنواع الأسمدة . أنواع أنظمة التغذية للتسمين . اسم المعالجات . والهدف من مثل هذه التجارب المقارنة بين متوسطات المجتمعات المناظرة للأربع أنواع من المعالجات .

ومن مميزات جمع البيانات وفقاً لتصميم تجريبي أنه يمكننا من الحصول على قدر من المعلومات أكبر من ذلك الذي يتم الحصول عليه إذا

لم تكن البيانات قد جمعت وفقاً لتصميم تجريبي . كما أنه يمكن البساحث من تحليل البيانات بأسلوب بسيط يعرف بتحليل التباين .

وموف نتناول في هذا الفصل تحليل التباين لتصميم تام العشوائية ،
ولتصميم القطاعات الكاملة العشوائية .

(٥ - ٢) المقارنة بين أكثر من متوسطي مجتمعين :

التصميم التام العشوائية :

Completely randomized design:

التصميم التام العشوائية هو ذلك التصميم الذي يتم فيه اختيار عينات عشوائية مستقلة من كل مجتمع من المجتمعات تحت الدراسة . وباستخدام هذه العينات يمكن المقارنة بين متوسطات هذه المجتمعات ، ويوضح الجدول (٥ - ١) الرموز المتعلقة بالتصميم التام العشوائية وذلك على فرض أننا نريد المقارنة بين متوسطات ل من المجتمعات .

وتتم المقارنة بين متوسطات المجتمعات (المعالجات) $\mu_1, \mu_2, \dots, \mu_L$ ،
لتقرير ما إذا كان هناك فرق بينها من خلال دراسة الاختلاف (التباين)
بين الأوساط الحسابية للعينات . فالإختلاف الأكبر دليل أقوى على وجود فرق
بين $\mu_1, \mu_2, \dots, \mu_L$ ، ولقياس هذا الاختلاف بالمجموع المرجح لمربعات
انحرافات الأوساط الحسابية للعينات $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_L$ عن الوسط العام
 \bar{y} ويعرف هذا المجموع بمجموع المربعات بين متوسطات العينات أو مجموع
المربعات بين المعالجات . ويمكن التعبير عنه كالآتي :

$$\text{مجموع المربعات بين المعالجات} = \sum_{j=1}^L n_j (\bar{y}_j - \bar{y})^2$$

جدول (١-٥)

الرموز المستخدمة في التصميم التام العشوائية

المجموعات (المعالجات)	
$1 \quad 2 \quad 3 \dots L$ $\mu_1 \quad \mu_2 \quad \mu_3 \dots \mu_L$ $\sigma_1^2 \quad \sigma_2^2 \quad \sigma_3^2 \dots \sigma_L^2$	<p>المتوسط التباين</p>
العينات العشوائية المستقلة	
$1 \quad 2 \quad 3 \dots L$ $n_1 \quad n_2 \quad n_3 \dots n_L$ $m_1 \quad m_2 \quad m_3 \dots m_L$ $\bar{m}_1 \quad \bar{m}_2 \quad \bar{m}_3 \dots \bar{m}_L$	
<p>حجم العينة</p> <p>مجموع المشاهدات</p> <p>الوسط الحسابي للعينة</p>	
<p>العدد الكلي للمشاهدات = $n = n_1 + n_2 + \dots + n_L$</p> <p>مجموع n من المشاهدات = $\sum_{j=1}^n \frac{n_j}{1-j}$ م</p> <p>الوسط العام = \bar{m}</p> <p>مجموع مربعات n من المشاهدات = $\sum_{j=1}^n \frac{n_j}{1-j} m_j^2$</p>	

ويلاحظ أنه تم ترجيح مربع كل فرق بين الوسط الحسابي للعينة والوسط العام بحجم العينة المناظرة كما يلاحظ أيضاً أن هذا المجموع سيكون كبيراً إذا كان الاختلاف كبيراً بين الأوساط الحسابية للعينات . فيفرض أننا نريد اختبار فرض العدم بتساوي متوسطات L من المعالجات ، أي أن :

$$\mu_1 = \mu_2 = \dots = \mu_L = \mu_0$$

في مقابل الفرض البديل :

H_1 : يوجد فرق بين متوسطين على الأقل .

في هذه الحالة نجد أن القيم الكبيرة لمجموع المربعات بين المعالجات هي التي ستؤيد الفرض البديل . بمعنى أنه إذا كان مجموع مربعات الفروق بين الأوساط الحسابية للعينات والوسط العام كبيراً فسيكون هناك اتجاه لتأييد الفرض بوجود فرق بين متوسطات المجتمعات .

والآن نتساءل إلى أي مدى يمكن اعتبار مجموع المربعات بين المعالجات كبيراً قبل رفض فرض العدم وقبول البديل ؟ نتوقف الإجابة على مقارنة هذا المجموع كقياس للاختلاف بين متوسطات العينات بمقياس آخر للاختلاف داخل العينات .

ويمكن قياس الاختلاف داخل العينات بإيجاد مجموع مشترك لمجاميع مربعات انحرافات المشاهدات داخل العينة عن وسطها الحسابي ويعرف هذا المجموع بمجموع المربعات داخل العينات أو ، بمجموع مربعات الخطأ حيث أنه يقيس الاختلاف الغير مفسر بواسطة الفروق بين متوسطات العينات ويمكن التعبير عن هذا المجموع بالصورة الآتية :

$$\text{مجموع مربعات الخطأ} = \sum_{j=1}^n (s_{1j} - \bar{s}_1)^2 + \sum_{j=1}^n (s_{2j} - \bar{s}_2)^2 + \dots + \sum_{j=1}^n (s_{rj} - \bar{s}_r)^2$$

$$+ \dots + \sum_{j=1}^n (s_{rj} - \bar{s}_r)^2$$

$$= \sum_{j=1}^n \sum_{r=1}^r (s_{rj} - \bar{s}_r)^2$$

حيث تشير s_{rj} إلى الملاحظة رقم j في العينة رقم r

ولمقارنة الاختلاف بين متوسطات العينات بالاختلاف داخل العينات
نتبع ما يلي :

١ - إيجاد متوسط المربعات بين المعالجات بقسمة مجموع المربعات بين المعالجات على درجات الحرية المرتبطة به (ل - ١) وهي عبارة عن درجة واحدة لمتوسط كل معالجة من ل من المعالجات مطروحاً منها الواحد الصحيح حيث تم تقدير الوسط العام أي أن :

$$\text{متوسط المربعات بين المعالجات} = \frac{\text{مجموع المربعات بين المعالجات}}{\text{ل - ١}}$$

٢ - إيجاد متوسط مربعات الخطأ بقسمة مجموع مربعات الخطأ على درجات الحرية المرتبطة به (ن - ل) وهي عبارة عن درجة لكل مشاهدة من ن من المشاهدات مطروحاً منها ل (عدد متوسطات العينات التي يتم تقديرها) :

$$\text{متوسط مربعات الخطأ} = \frac{\text{مجموع مربعات الخطأ}}{\text{ن - ل}}$$

٣ - حساب إحصائية ف :

$$\text{ف} = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

حيث تشير القيم الكبيرة لإحصائية ف إلى وجود فروق كبيرة بين متوسطات العينات وبالتالي تأييد الفرض البديل بوجود فرق بين متوسطات المجتمعات . ويلاحظ أن إحصائية ف تتضمن المقارنة بين مصدرين للاختلاف (التباين) أحدهما يرجع إلى الفروق بين متوسطات العينات والآخر يرجع إلى الفروق داخل العينات ولذلك سُمي هذا الأسلوب لمقارنة متوسطات المجتمعات بتحليل التباين (ANOVA) Analysis of Variance .

ويمكن تلخيص عناصر اختبار الفرق بين متوسطات ل من المجتمعات والشروط الواجب توافرها كالآتي :

اختبار الفرق بين متوسطات ل من المعالجات باستخدام التصميم التام العشوائية

الفروض الإحصائية :

$$\sigma^2_H : \mu_1 = \mu_2 = \dots = \mu_k = \dots = \mu_n$$

σ^2_H : يوجد فرق بين متوسطين على الأقل

إحصائية الاختبار : $F = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$

الشروط : ١ - يتبع كل مجتمع من المجتمعات المراد المقارنة بين متوسطاتها لتوزيع معتدل

٢ - تباينات المجتمعات متساوية .

٣ - العينات عشوائية ومستقلة .

منطقة الرفض : $F > F_{\alpha}(n-1, k-1)$

حيث $n-1 = l-1 =$ عدد درجات الحرية المرتبطة بمتوسط المربعات بين المعالجات ، $n-1 =$ عدد درجات الحرية المرتبطة بمتوسط مربعات الخطأ .

وبالرغم من إمكانية استخدام الصيغ السابقة لحساب مجموع المربعات بين المعالجات ومجموع مربعات الخطأ إلا أنه يوجد صيغ أخرى أكثر بساطة في الحساب . بالإضافة إلى ذلك إمكانية اللجوء إلى فكرة تجزئة مجموع مربعات انحرافات كل المشاهدات عن الوسط العام ويعبرف هذا المجموع بمجموع المربعات الكلي . أي أن :

$$\text{مجموع المربعات الكلي} = \sum_{i=1}^n (x_i - \bar{x})^2$$

= مجموع المربعات بين المعالجات + مجموع مربعات الخطأ

والفائدة التي تعود علينا من وراء فكرة هذه التجزئة هي إمكانية إيجاد مجموع مربعات الخطأ كمتعم حسابي أي أن :

مجموع مربعات الخطأ = مجموع المربعات الكلي - مجموع المربعات بين المعالجات

صيغ الحسابات اللازمة لتحليل التباين للتصميم التام العشوائية

$$\text{معامل التصحيح} = \frac{(\text{مجموع كل المشاهدات})^2}{\text{العدد الكلي للمشاهدات}}$$

$$= \frac{(\text{مجموع من } y^2)}{n}$$

مجموع المربعات الكلي = مجموع مربعات المشاهدات - معامل التصحيح

$$= \text{مجموع من } y^2 - \text{معامل التصحيح}$$

$$\text{مجموع المربعات بين المعالجات} = \left(\begin{array}{l} \text{مجموع مربعات مجاميع المعالجات} \\ \text{مقسوماً كل منها على عدد} \\ \text{المشاهدات للمعالجة} \end{array} \right) - \text{معامل التصحيح}$$

$$= \frac{y_1^2}{n_1} + \frac{y_2^2}{n_2} + \dots + \frac{y_j^2}{n_j} - \text{معامل التصحيح}$$

مجموع مربعات الخطأ = مجموع المربعات الكلي - مجموع المربعات بين المعالجات

$$\text{متوسط المربعات بين المعالجات} = \frac{\text{مجموع المربعات بين المعالجات}}{j - 1}$$

$$\text{متوسط مربعات الخطأ} = \frac{\text{مجموع مربعات الخطأ}}{n - j}$$

$$\text{إحصائية الاختبار : F} = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

وعادة ما يتم تلخيص نتائج الحسابات في صورة جدول يعرف بجدول تحليل التباين . والجدول (٥ - ٢) يوضح جدول تطليل التباين لتصميم تام العشوائية . ويوضح فيه مصدر الاختلاف ، درجات الحرية ، مجموع المربعات ، متوسط المربعات وإحصائية الاختبار ف .

جدول (٥ - ٢)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات	ل - ١	مجموع المربعات بين المعالجات	متوسط مربعات المعالجات	متوسط مربعات المعالجات
الخطأ	ل - ن	مجموع مربعات الخطأ	متوسط مربعات الخطأ	متوسط مربعات الخطأ
الكل	ن - ١	مجموع مربعات الكل		

مثال (١) :

أرادت إحدى المؤسسات الائتمانية مقارنة متوسطات الديون المستحقة الدفع على العملاء في ثلاث مستويات مختلفة للدخل السنوي بالجنهيات (أقل من ١٢٠٠٠ ، ١٢٠٠٠ - ٢٥٠٠٠ ، أكبر من ٢٥٠٠٠) فقامت باختيار عينة مكونة من حسابات عشرة عملاء من كل مجموعة دخلية ، وتم تسجيل الديون المستحقة الدفع على كل عميل وجدول (٥ - ٣) يلخص النتائج التي تم الحصول عليها . المطلوب استخدام هذه البيانات لاختبار الفرض بأن متوسطات الديون المستحقة الدفع في المجموعات الدخلية الثلاثة متساوية في مقابل الفرض البديل بأنها مختلفة . استخدم مستوى معنوية ٠,٠٥ .

جدول رقم (٥ - ٣)

الديون المستحقة الدفع في المجموعات الدخلية المختلفة

أقل من ١٢٠٠٠	١٢٠٠٠ - ٢٥٠٠٠	أكثر ٢٥٠٠٠
١٤٨	٥١٣	٣٣٥
٧٦	٢٦٤	٦٤٣
٣٩٣	٤٣٣	٢١٦
٥٢٠	٩٤	٥٣٦
٢٣٦	٥٣٥	١٢٨
١٣٤	٣٢٧	٧٢٣
٥٥	٢١٤	٢٥٨
١٦٦	١٣٥	٣٨٠
٤١٥	٢٨٠	٥٩٤
١٥٣	٣٠٤	٤٦٥
٢٢٩٦	٣٠٩٩	٤٢٧٨

الحل :

حيث أننا نريد أن نختبر فرض العدم بتساوي متوسطات الديون المستحقة الدفع في المجموعات الدخلية الثلاثة ، وبفرض أن μ_1 ، μ_2 ، μ_3 تمثل على الترتيب متوسط الديون المستحقة الدفع في المجموعة الدخلية المنخفضة والمتوسطة والمرتفعة المستوى فتكون عناصر الاختبار كالاتي :

الفروض الإحصائية :

$$\mu_1 = \mu_2 = \mu_3 : H_0$$

H_1 : يوجد فرق بين متوسطين على الأقل

$$\text{إحصائية الاختبار : } F = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

الشروط : ١ - تتوزع الديون المستحقة الدفع في كل مجموعة دخلية توزيعاً معتدلاً .

٢ - تساوي التباين لتوزيعات الديون المستحقة للدفع في المجموعات الدخلية الثلاثة .

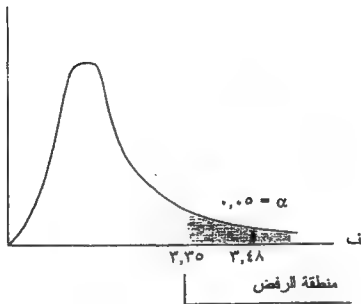
٣ - العينات عشوائية ومستقلة عن بعضها البعض .

منطقة الرفض : $F > F_{(27, 17, 0.05)}$

$$\text{حيث } 17 = 1 - 3 = 1 - \text{د.ج} = 17$$

$$27 = 3 - 30 = \text{د.ن} - 30 = 27$$

$$F_{(27, 17, 0.05)} = F_{(27, 17, 0.05)} = 3.35 \quad (\text{أنظر الشكل ٥ - ١})$$



شكل (٥ - ١)

من جدول (٥ - ٣) نجد أن :

$$٢٢٩٦ = ١م ، ٣٠٩٩ = ٢م ، ٤٢٧٨ = ٣م$$

$$٩٦٧٣ = ٣م + ٢م + ١م$$

$$مجمس = ١(١٤٨) + ٢(٧٦) + + ٣(٤٦٥) = ٤٠٨٨٣٤١$$

$$معامل التصحيح = \frac{مجمس}{ن} = \frac{٩٦٧٣}{٣٠} = ٣١١٨٨٩٧,٦$$

مجموع المربعات الكلي = مجمس - معامل التصحيح

$$= ٣١١٨٨٩٧,٦ - ٤٠٨٨٣٤١ =$$

$$٩٦٩٤٤٣,٤ =$$

$$مجموع المربعات بين المعالجات = \frac{١م}{١ن} + \frac{٢م}{٢ن} + \frac{٣م}{٣ن} - معامل التصحيح$$

$$= \frac{٢(٢٢٩٦)}{١٠} + \frac{٣(٣٠٩٩)}{١٠} + \frac{٤(٤٢٧٨)}{١٠} - ٣١١٨٨٩٧,٦ =$$

$$= ١٩٨٧٧٢,٥ - ٣٣١٧٦٧,١ =$$

مجموع مربعات الخطأ = مجموع المربعات الكلي - مجموع المربعات بين المعالجات

$$= ٩٦٩٤٤٣,٤ - ١٩٨٧٧٢,٥ = ٧٧٠٦٧٠,٩$$

$$متوسط المربعات بين المعالجات = \frac{مجموع المربعات بين المعالجات}{١ - ج}$$

$$= \frac{١٩٨٧٧٢,٥}{٢} = ٩٩٣٨٦,٢٥$$

$$متوسط مربعات الخطأ = \frac{مجموع مربعات الخطأ}{ن - ج}$$

$$= \frac{٧٧٠٦٧٠,٩}{٢٧} = ٢٨٥٤٣,٤$$

$$ف = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

$$3,48 = \frac{93386,25}{28543,4} =$$

وحيث أن قيمة ف المحسوبة تقع في منطقة الرفض كما هو مبين بالشكل (٥ - ١) فإننا نستنتج أن هناك اختلاف بين متوسط الديون المستحقة الدفع في مجموعتين على الأقل من المجموعات الدخلية وذلك عند مستوى معنوية ٠,٠٥ وجدول (٥ - ٤) يوضح جدول تحليل التباين لمثالنا الحالي :

جدول (٥ - ٤)

جدول تحليل التباين لمثال (١)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات	٢	١٩٨٧٧٢,٥	٩٩٣٨٦,٢٥	٣,٤٨
الخطأ	٢٧	٧٧٠٦٧٠,٩	٢٨٥٤٣,٤	
الكلي	٢٩	٩٦٩٤٤٣,٤		

ونظراً لما يتضمنه التصميم التام العشوائية من اختيار لعينات عشوائية مستقلة فإنه يمكن إيجاد فترة ثقة لمتوسط المعالجة الواحدة وكذلك للفرق بين متوسطي معالجتين مع الأخذ في الاعتبار أن يستخدم متوسط مربعات الخطأ كمقدر للتباين σ^2 أي أن

$$ع^2 = \text{متوسط مربعات الخطأ} = \frac{\text{مجموع مربعات الخطأ}}{ن - ١}$$

وبناء على ذلك يمكن تلخيص كيفية إيجاد فترة ثقة لمتوسط المعالجة
الواحدة وكذلك للفرق بين متوسطي معالجتي كالآتي :

فترة ثقة ١٠٠ (١ - α) % لمتوسط للمعالجة ر
$\bar{r} \pm t_{(n-1, \frac{\alpha}{2})} \frac{e}{\sqrt{n}}$
فترة ثقة ١٠٠ (١ - α) % للفرق بين متوسطي المعالجتين ر ، ز
$(\bar{r} - \bar{z}) \pm t_{(n-1, \frac{\alpha}{2})} e \sqrt{\frac{1}{n} + \frac{1}{n_z}}$
حيث : $e = \sqrt{\text{متوسط مربعات الخطأ}}$

مثال (٢) :

في المثال (١) المطلوب إيجاد فترة ثقة ٩٥% لمتوسط الديون
المستحقة الدفع لمجموعة العملاء الذين يقل دخلهم السنوي عن ١٢٠٠٠ جنيها .

الحل :

من جدول تحليل التباين (٥ - ٤) نجد أن متوسط مربعات الخطأ =
٢٨٥٤٣,٤ وبالتالي فإن :

$$e = \sqrt{\text{متوسط مربعات الخطأ}}$$

$$168,9 = \sqrt{28543,4}$$

كما يمكن إيجاد الوسط الحسابي لعينة الديون المستحقة الدفع لمجموعة
العملاء الذين يقل دخلهم عن ١٢٠٠٠ جنيها كالآتي :

$$\overline{س} = \frac{١٦}{١٠} = \frac{٢٢٩٦}{٢٢٩,٦}$$

وبناء على ذلك تكون فترة ثقة ٩٥٪ لمتوسط الديون المستحقة الدفع لمجموعة العملاء ، الذين يقل دخلهم عن ١٢٠٠٠ جنيها على الصورة :

$$\begin{aligned} & \overline{س} \pm t \left(\frac{\sigma}{\sqrt{n}} \right) \\ & \overline{س} \pm t (27.0.20) = \\ & = 229.6 \pm 2.052 \left(\frac{168.9}{\sqrt{10}} \right) = \\ & = 229.6 \pm 109.6 = \\ & = (120 , 339.2) \end{aligned}$$

ومن الملاحظ أن هذه الفترة تتسم بالاتساع ويرجع السبب في ذلك إلى زيادة درجة التفاوت بين مفردات العينة ، فهي تتفاوت من ٥٥ جنيها إلى ٢٥٠ جنيها ولكن إذا أردنا الحصول على تقدير دقيق لمتوسط المعالجة بفترة ثقة أضيق من تلك الفترة فيجب تزويد حجم العينة .

مثال (٣) :

في المثال (١) أوجد فترة ثقة ٩٥٪ للفرق بين متوسطي الديون المستحقة الدفع على العملاء الذين يقل دخلهم السنوي عن ١٢٠٠٠ جنيها والذين يزيد دخلهم السنوي عن ٢٥٠٠٠ جنيها .

الحل :

من المثال (١) يمكن إيجاد الوسط الحسابي لعينة الديون المستحقة للدفع على العملاء الذين يزيد دخلهم عن ٢٥٠٠٠ جنيها في السنة كالآتي :

$$٤٢٧,٨ = \frac{٤٢٧٨}{١٠} = \frac{٢٣}{٣٨} = \overline{٣٨}$$

ومن المثال (٢) وجدنا أن $\overline{١٠} = ٢٢٩,٦$. وبالتالي تكون فترة تق
٩٥٪ للفرق ($\overline{١٢} - \overline{١١}$) على الصورة :

$$\begin{aligned} & \sqrt{\frac{1}{10} + \frac{1}{38}} \sqrt{E\left(1 - 0,05 \cdot \frac{\alpha}{2}\right)} \pm (\overline{10} - \overline{38}) \\ & \sqrt{\frac{1}{10} + \frac{1}{10}} \sqrt{(168,9)(2,052) \pm (229,6 - 427,8)} = \\ & 100,0 \pm 198,2 = \\ & = (303,2, 427,2) \end{aligned}$$

ومن الملاحظ أن هذه الفترة للفرق بين المتوسطين ($\overline{١٢} - \overline{١١}$) تـ
بالإتساع الشديد كما هو الحال بالنسبة لفترة الثقة للمتوسط الواحد ويرجع السبـ
في ذلك أيضا إلى التفاوت الكبير بين المفردات داخل العينة الواحدة. ولذلك حتـ
نستطيع الحصول على فترة أضيق لابد من تزويد حجم العينات الثلاثة
ويلاحظ أيضا أن هذه الفترة تحتوي على قيم موجبة فقط مما يجعلنا نتق بدرـ
٩٥٪ في أن يكون متوسط الديون المستحقة الدفع على العملاء انين يزيد دـ
السنوي عن ٢٥٠٠٠ جنيهها أعلى من نظيره بالنسبة للعملاء الذين يقل دخلـ
السنوي عن ١٢٠٠٠ جنيهها .

(٥ - ٣) تصميم القطاعات الكاملة العشوائية :

Completely Randomized Block Design:

يستخدم تصميم القطاعات الكاملة العشوائية مجموعات من الوحدات التجريبية المتجانسة للمقارنة بين متوسطات المجتمعات المتعلقة بعدد من المعالجات . ففي هذا النوع من التصميمات يتم تقسيم الوحدات التجريبية إلى مجموعات بشرط أن تكون هذه الوحدات داخل المجموعة الواحدة متجانسة ومساوية في عددها لعدد المعالجات . كما يشترط أن يكون توزيع المعالجات داخل المجموعة عشوائيا . ويطلق على كل مجموعة من هذه المجموعات اسم قطاع BLOCK . ويوضح الجدول (٥ - ٥) الرموز التي ستستخدم عند تحليل نتائج تصميم القطاعات الكاملة العشوائية . وذلك بفرض أن لدينا ط من القطاعات وأن عدد المعالجات المطلوب المقارنة بين متوسطاتها يساوي ل .

جدول (٥ - ٥)

الرموز المستخدمة في تصميم القطاعات الكاملة العشوائية

المعالجة	المعالجة	المعالجة	
(١)	(٢)	المعالجة (ل)
ط	ط	ط
١٢	٢٢	مجموع المشاهدات
قطاع	قطاع	قطاع	قطاع
(١)	(٢)	(ط)	المعالجة (ط)
ل	ل	ل
ل	ل	ل
١٢	٢٢	مجموع المشاهدات
العدد الكلي للملاحظات = ل ط = ن			
مجموع ن من المشاهدات = ج س			
مجموع ن من مربعات المشاهدات = ج س ^٢			

وهدفنا الآن هو استخدام تصميم القطاعات الكاملة العشوائية لاختبار فرض العدم بتساوي متوسطات المعالجات في مقابل البديل بوجود فرق بين هذه المتوسطات . أي أن :

$$\mu_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

H_1 : يوجد اختلاف بين متوسطي معالجتين على الأقل

ونلاحظ أن هذه الفروض هي نفس الفروض التي سبق اختبارها في التصميم التام العشوائية . وكذلك تأخذ إحصائية الاختبار نفس الشكل الذي سبق استخدامه في التصميم التام العشوائية :

$$\text{إحصائية الاختبار : } F = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

حيث يحسب البسط (متوسط المربعات بين المعالجات) بنفس الطريقة التي حسب بها في التصميم التام العشوائية في حين يحسب المقام (متوسط مربعات الخطأ) بطريقة مختلفة . حيث يجرى مجموع المربعات الكلي في تصميم القطاعات الكاملة العشوائية إلى ثلاثة أجزاء بدلا من جزئين . حيث أن :

$$\text{مجموع المربعات الكلي} = \text{مجموع المربعات بين المعالجات} + \text{مجموع المربعات بين القطاعات} + \text{مجموع مربعات الخطأ}$$

وبالتالي فإن :

$$\text{مجموع مربعات الخطأ} = \text{مجموع المربعات الكلي} - \text{مجموع المربعات بين المعالجات} - \text{مجموع المربعات بين القطاعات}$$

وهذا يعني أن مجموع مربعات الخطأ في تصميم القطاعات الكاملة العشوائية يساوي مجموع مربعات الخطأ في التصميم التام العشوائية مطروحا

منه مجموع المربعات بين القطاعات . وبناءا على ذلك نستطيع القول أن تصميم القطاعات الكاملة العشوائية يسمح بإزالة الاختلاف بين القطاعات من الاختلاف داخل العينات . وهو من شأنه تقليل متوسط مربعات الخطأ (والذي يمثل المقام في إحصائية ف) وبالتالي تكون هناك فرصة أكبر لملاحظة الفروق بين متوسطات المعالجات . ويمكن تلخيص عناصر اختبار الفرق بين متوسطات المعالجات باستخدام القطاعات الكاملة العشوائية وكذلك الشروط الواجب توافرها كالاتي :

لختبار الفرق بين متوسطات ل من المعالجات باستخدام تصميم القطاعات الكاملة العشوائية

الفروض الإحصائية :

$$\sigma^2_H : \mu_1 = \mu_2 = \dots = \mu_L$$

σ^2_H : يوجد اختلاف بين متوسطي معالجتين على الأقل

$$\text{إحصائية الاختبار : } F = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

لتشروط : ١ - التوزيعات الاحتمالية للمشاهدات المناظرة لكل توليفات

القطاعات مع المعالجات معتدلة .

٢ - تباينات التوزيعات الاحتمالية متساوية .

منطقة الرفض : $F > F(\alpha, \nu, \nu)$

$$\text{حيث } \nu = L - 1, \quad \nu = N - P - L + 1$$

ويمكن تلخيص الصيغ اللازمة لتحليل تصميم القطاعات الكاملة العشوائية في الآتي :

صيغ الحسابات اللازمة لتحليل تصميم القطاعات الكاملة العشوائية

$$\text{معامل التصحيح} = \frac{(\text{مجموع كل المشاهدات})^2}{\text{العدد الكلي للمشاهدات}}$$

$$= \frac{\left(\sum_{j=1}^n \sum_{i=1}^m x_{ij} \right)^2}{N}$$

مجموع المربعات الكلي = مجموع مربعات كل المشاهدات - معامل التصحيح

$$= \sum_{j=1}^n \sum_{i=1}^m x_{ij}^2 - \text{معامل التصحيح}$$

$$\text{مجموع المربعات بين المعالجات} = \left(\begin{array}{l} \text{مجموع مربعات مجاميع المعالجات} \\ \text{مقسوماً كل منها على عدد} \\ \text{المشاهدات للمعالجة} \end{array} \right) - \text{معامل التصحيح}$$

$$= \frac{x_{11}^2}{p} + \dots + \frac{x_{1m}^2}{p} + \frac{x_{21}^2}{p} + \dots + \frac{x_{2m}^2}{p} - \text{معامل التصحيح}$$

$$\text{مجموع المربعات بين القطاعات} = \left(\begin{array}{l} \text{مجموع مربعات مجاميع القطاعات} \\ \text{مقسوماً كل منها على عدد} \\ \text{المشاهدات للقطاع} \end{array} \right) - \text{معامل التصحيح}$$

$$= \frac{x_{11}^2}{q} + \dots + \frac{x_{1m}^2}{q} + \frac{x_{21}^2}{q} + \dots + \frac{x_{2m}^2}{q} - \text{معامل التصحيح}$$

مجموع مربعات الخطأ = مجموع المربعات الكلي - مجموع المربعات بين المعالجات

- مجموع المربعات بين القطاعات

$$\text{متوسط المربعات بين المعالجات} = \frac{\text{مجموع المربعات بين المعالجات}}{1 - L}$$

$$\frac{\text{مجموع المربعات بين القطاعات}}{\text{ط} - ١} = \text{متوسط المربعات بين القطاعات}$$

$$\frac{\text{مجموع مربعات الخطأ}}{\text{ن} - \text{ل} - \text{ط} - ١} = \text{متوسط مربعات الخطأ}$$

$$\frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}} = \text{إحصائية الاختبار : ف}$$

وجداول (٥ - ٦) يوضح تحليل التباين لتصميم القطاعات الكاملة

العشوائية :

جدول (٥ - ٦)

جدول تحليل التباين لتصميم القطاعات الكاملة العشوائية

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات	ل - ١	بين المعالجات	بين المعالجات	متوسط مربعات بين المعالجات
القطاعات	ط - ١	بين القطاعات	بين القطاعات	متوسط مربعات بين القطاعات
الخطأ	ن - ل - ط - ١	الخطأ	الخطأ	متوسط مربعات الخطأ
الكلية	ن - ١	الكلية		

مثال (٤) :

أرادت إحدى الشركات الصناعية الكبرى المنتجة للملابس الجاهزة القيام بتجربة لدراسة تأثير زيادة الأجر / الساعة على إنتاجية العاملين بها فقامت باستخدام أربعة أنظمة للدفع (معالجات)

المعالجة (١) : عدم زيادة الأجر / الساعة .

المعالجة (٢) : زيادة الأجر / الساعة بمقدار ٥٠ قرشا .

المعالجة (٣) : زيادة الأجر / الساعة بمقدار ١٠٠ قرشا .

المعالجة (٤) : زيادة الأجر / الساعة بمقدار ١٥٠ قرشا .

وقامت باختيار اثني عشر عاملا ، تم تقسيمهم إلى ثلاثة مجموعات على أساس المدة التي قضاها العامل بالشركة على أن يكون بكل مجموعة أربعة عمال توزع عليهم الأربعة معالجات عشوائيا . وقد تم متابعة هؤلاء العمال لمدة ثلاثة أسابيع وقياس إنتاجيتهم بناءا على متوسط عدد الوحدات السليمة المنتجة في الساعة . وجدول (٥ - ٧) يلخص النتائج التي تم الحصول عليها . والمطلوب استخدام تحليل التباين لتحديد ما إذا كانت البيانات تدل على وجود فرق بين متوسطات إنتاجية العمال في ظل الأنظمة الأربعة للدفع (استخدم مستوى معنوية ٠,٠٥) .

جدول (٥ - ٧)

متوسط إنتاجية العامل في الساعة

المجموع					المعالجات
	(٤)	(٣)	(٢)	(١)	المدة التي قضاها العامل بالشركة
١١,٧	٣,٢	٣,١	٣,٠	٢,٤	أقل من سنة
٢٢,٣	٥,٧	٥,٩	٦,١	٤,٦	١ - ٥ سنوات
٢٦,٦	٧,٣	٧,٢	٧,٠	٥,١	أكثر من خمس سنوات
٦٠,٦	١٦,٢	١٦,٢	١٦,١	١٢,١	المجموع

الحل :

بفرض أن $1\mu, 2\mu, 3\mu, 4\mu$ تمثل على الترتيب متوسط إنتاجية العامل في ظل النظام الأول ، الثاني ، الثالث ، الرابع لدفع الأجر فتكون عناصر اختبار تحليل التباين كالاتي :

الفروض الإحصائية :

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$$

H_1 : يوجد فرق بين متوسطين على الأقل

إحصائية الاختبار : $F = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$

الشروط : ١ - التوزيعات الاحتمالية لمتوسط عدد الوحدات المنتجة في الساعة المناظرة لكل توليفات المدة التي قضاها العامل في الشركة ونظام دفع الأجر معتدلة .

٢ - تباينات التوزيعات الاحتمالية متساوية .

منطقة الرفض : $F > F_{(1, 3, 0.05)}$

حيث أن : $1 - \alpha = 1 - 0.05 = 0.95$ ،

$$1 - \alpha = 0.95 = 1 - 0.05 = 0.95$$

$$\therefore F_{(1, 3, 0.05)} = F_{(1, 3, 0.05)} = 10.13$$

وباستخدام بيانات الجدول (٥ - ٧) نجد أن :

$$MSB = 2(4, 2) + (3) + \dots + (7, 3) = 3.042$$

$$\frac{\sum (\text{مـ سـ})^2}{\text{ن}} = \text{معامل التصحيح}$$

$$306,03 = \frac{\sum (60,6)^2}{12} =$$

مجموع المربعات الكلية = مـ سـ² - معامل التصحيح

$$34,39 = 306,03 - 340,42 =$$

$$\text{مجموع المربعات بين المعالجات} = \frac{\sum 1^4}{\text{ط}} + \frac{\sum 2^4}{\text{ط}} + \frac{\sum 3^4}{\text{ط}} + \frac{\sum 4^4}{\text{ط}} - \text{معامل التصحيح}$$

$$306,03 - \frac{\sum (16,2)^2}{3} + \frac{\sum (16,2)^2}{3} + \frac{\sum (16,1)^2}{3} + \frac{\sum (12,1)^2}{3} =$$

$$4,14 = 306,03 - 310,17 =$$

$$\text{مجموع المربعات بين القطاعات} = \frac{\sum 1^4}{\text{ل}} + \frac{\sum 2^4}{\text{ل}} + \frac{\sum 3^4}{\text{ل}} - \text{معامل التصحيح}$$

$$306,03 - \frac{\sum (26,6)^2}{4} + \frac{\sum (22,3)^2}{4} + \frac{\sum (11,7)^2}{4} =$$

$$29,41 = 306,03 - 335,44 =$$

مجموع مربعات الخطأ¹ = مجموع المربعات الكلية - مجموع المربعات بين المعالجات - مجموع المربعات بين القطاعات

$$0,84 = 29,41 - 4,14 - 34,39 =$$

$$\frac{\text{مجموع المربعات بين المعالجات}}{1 - \text{ل}} = \text{متوسط المربعات بين المعالجات}$$

$$1,38 = \frac{4,14}{1 - 4} =$$

$$\frac{\text{مجموع المربعات بين القطاعات}}{\text{ط} - 1} = \text{متوسط المربعات بين القطاعات}$$

$$14.71 = \frac{29.41}{1-3} =$$

$$\frac{\text{مجموع مربعات الخطأ}}{\text{ن} - \text{ل} - \text{ط} + 1} = \text{متوسط مربعات الخطأ}$$

$$0.14 = \frac{0.84}{1+3-4-12} =$$

$$\frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}} = \text{ف}$$

$$9.86 = \frac{1.38}{0.14} =$$

وحيث أن قيمة ف المحسوبة (٩,٨٦) أكبر من قيمة ف الجدولية (٤,٧٦) فأنا نرفض فرض العدم بمتساوي المتوسطات ونقبل البديل بوجود اختلاف بين متوسطين على الأقل وذلك عند مستوى معنوية ٠,٠٥ و جدول (٥ - ٨) يوضح جدول تحليل التباين لمثالنا الحالي :

جدول (٥ - ٨)

جدول تحليل التباين لمثال (٤)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات	٣	٤,١٤	١,٣٨	٩,٨٦
القطاعات	٢	٢٩,٤١	١٤,٧١	
الخطأ	٦	٠,٨٤	٠,١٤	
الكلي	١١	٣٤,٣٩		

ويمكن إيجاد فترة ثقة للفرق بين متوسطي أي معالجتين مع مراعاة استخدام متوسط مربعات الخطأ كمقدر للتباين σ^2 . أي أن :

$$\begin{aligned} \text{ع}^2 &= \text{متوسط مربعات الخطأ} \\ &= \frac{\text{مجموع مربعات الخطأ}}{ن - ل - ط + ١} \end{aligned}$$

وبناء على ذلك تكون صيغة فترة الثقة كالآتي :

<p>فترة ثقة ١٠٠ (١ - α) % للفرق بين متوسطي معالجتين (لـ - لـ)</p> $\begin{aligned} &(\bar{مـ} - \bar{مـ}) \pm t_{(١-\alpha, ٠.٥-ل-ط+١, ن)} \sqrt{\text{ع} \left(\frac{١}{ط} + \frac{١}{ط} \right)} \\ &= (\bar{مـ} - \bar{مـ}) \pm t_{(١-\alpha, ٠.٥-ل-ط+١, ن)} \sqrt{\text{ع} \left(\frac{٢}{ط} \right)} \end{aligned}$ <p>حيث : $\text{ع} = \sqrt{\text{متوسط مربعات الخطأ}}$</p>
--

مثال (٥) :

في المثال (٤) المطلوب إيجاد فترة ثقة ٩٠ % للفرق في متوسط الإنتاجية للمعالجات (١) ، (٢)

الحل :

من المثال (٤) نجد أن الوسط الحسابي لإنتاجية عينة العمال الذين لم يتغير أجرهم في الساعة (معالجة (١)) ولعينة العمال الذين زاد أجرهم بمقدار ٥٠ قرشا في الساعة (معالجة (٢)) هما على الترتيب :

$$\bar{م}_1 = \frac{12,1}{3} = 4,03$$

$$\bar{م}_2 = \frac{16,1}{3} = 5,37$$

من جدول تحليل التباين (٥ - ٨) نحصل على :

$$ع = \sqrt{\text{متوسط مربعات الخطأ}}$$

$$ع = 0,14 = \sqrt{0,37}$$

وبالتالي تكون فترة ثقة ٩٠٪ للفرق $0,14$ ، $0,14$ على الصورة :

$$(\bar{م}_2 - \bar{م}_1) \pm t \left(\frac{\alpha}{2}, n-d-1 \right) ع \sqrt{\frac{2}{ط}}$$

$$= (\bar{م}_2 - \bar{م}_1) \pm t (0,05, 1) ع \sqrt{\frac{2}{ط}}$$

$$= (5,37 - 4,03) \pm (1,943) (0,37) \sqrt{\frac{2}{3}}$$

$$= 1,34 \pm 0,59$$

$$= (-1,93 - 0,70)$$

وهذا يعني أن متوسط إنتاجية العامل بالنسبة للمعالجة (٢) يزيد عن نظيره بالنسبة للمعالجة (١) بمقدار يتراوح بين ٠,٧٥ ، ١,٩٣ .

وبالإضافة إلى ما سبق بشأن إجراء اختبار للفرق بين متوسطات المعالجات فإنه يمكن إجراء اختبار للفرق بين متوسطات القطاعات . فهذا الاختبار يمكننا من تقرير ما إذا كان تصميم القطاعات قد نجح في تخفيض

حجم الخطأ التجريبي أم لا . بمعنى أنه إذا كان هناك اختلاف بين متوسطات القطاعات فيكون ذلك دليلاً على أن الوحدات للتجريبية دخل القطاعات أكثر تجانساً منها بين القطاعات مما يبرر الحاجة لاستخدام تصميم القطاعات الكاملة العشوائية . وإجراء الاختبار لمتوسطات القطاعات قريب الشبه جداً من اختبار متوسطات المعالجات . حيث يتم مقارنة الاختلاف بين القطاعات ، مقاساً بمتوسط المربعات بين القطاعات ، بالاختلاف الناتج عن الخطأ التجريبي مقاساً بمتوسط مربعات الخطأ . ويمكن تلخيص هذا الاختبار في الآتي :

لختبار الفرق بين متوسطات ط من القطاعات باستخدام تصميم القطاعات الكاملة العشوائية

الفروض الإحصائية :

H_0 : متوسطات ط من القطاعات متساوية

H_1 : يوجد اختلاف بين متوسطي قطاعين على الأقل

إحصائية الاختبار : $F = \frac{\text{متوسط المربعات بين القطاعات}}{\text{متوسط مربعات الخطأ}}$

الشروط : نفس الشروط السابقة الذكر عند إجراء اختبار الفرق بين متوسطات المعالجات .

منطقة الرفض : $F > F_{(27, 170, 0.05)}$

حيث $17 = ط - 1$ ، $27 = ن - ل - ط + 1$

مثال (٦) :

في المثال (٤) والخاص باستخدام تصميم القطاعات الكاملة العشوائية لمقارنة متوسط إنتاجية العامل في ظل أربعة أنظمة لدفع الأجر . وكانت المدد المختلفة التي قضاها العامل بالشركة ممثلة للقطاعات . لختبر

الفرض بتساوي متوسط إنتاجية العامل في القطاعات الثلاثة . استخدم مستوى معنوية ٠,٠٥ .

الحل :

عناصر اختبار الفرق بين متوسطات القطاعات هي :

الفروض الإحصائية :

H_0 : متوسط إنتاجية العامل متساو في القطاعات الثلاثة .

H_1 : يوجد فرق بين متوسطين على الأقل

$$\text{إحصائية الاختبار : } F = \frac{\text{متوسط المربعات بين القطاعات}}{\text{متوسط مربعات الخطأ}}$$

الشروط : نفس الشروط للسابق ذكرها في حل المثال (٤)

منطقة الرفض : $F < F_{(٢٧, ١٧, ٠,٠٥)}$

حيث أن : $١٧ = ط - ١ - ٣ = ١ - ٢ = ٢$ ،

$$٢٧ = ن - ل - ط + ١ = ٦$$

$$\therefore F_{(٢٧, ١٧, ٠,٠٥)} = F_{(٦, ٢, ٠,٠٥)} = ٥,١٤$$

وفي المثال (٤) تم حساب متوسط المربعات بين القطاعات ، ومتوسط مربعات الخطأ وكانا على الترتيب ١٤,٧١ ، ١٤ وبالتعويض بهذه القيم في إحصائية F نحصل على :

$$F = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

$$105.07 = \frac{14.71}{0.14} =$$

وحيث أن قيمة F المحسوبة (105.07) تزيد بدرجة كبيرة عن قيمة F الجدولية (5.14) فإننا نستنتج أن البيانات كافية لتأييد الفرض البديل بوجود اختلاف في متوسط إنتاجية العامل في القطاعات الثلاثة الممتدة للمـ. التي قضاها العامل بالشركة . ومن ثم كان القرار باستخدام تصميم القطاعات الكاملة العشوائية قرارا حكيما . وكان لاستخدام هذا التصميم أثره الواضح في تخفيض مجموع مربعات الخطأ وبالتالي زيادة حجم المعلومات في التجربة وجداول (9 - 5) يوضح تحليل التباين الكامل لهذه التجربة .

جدول (5 - 9)

جدول تحليل التباين الكامل لمثال (6)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	F
المعالجات	3	4.14	1.38	9.86
القطاعات	2	29.41	14.71	105.07
الخطأ	6	0.84	0.14	
الكلي	11	34.39		

ونود أن نشير في نهاية هذا البند إلى ضرورة تفسير نتائج اختبار الفرق بين متوسطات القطاعات بشيء من الحيط والحذر وخاصة إذا كانت قيمة F المحسوبة لا تقع في منطقة الرفض . فهذا لا يعني بالضرورة أن تكون

متوسطات القطاعات متساوية وبالتالي لا تكون هناك فائدة من استخدام القطاعات. فالوصول إلى مثل هذا الاستنتاج يناظر القول بقبول فرض العدم والذي يجب تجنبه بسبب عدم معلومية احتمال الوقوع في خطأ من النوع الثاني (قبول فرض العدم وهو غير صحيح في الواقع) وبناءً على ذلك نستطيع القول أنه في حالة فشل الاختبار في الوصول إلى قرار قاطع بشأن الفروق بين متوسطات القطاعات فيمكن للقائم بالتجربة أن يستخدم تصميم للقطاعات الكاملة العشوائية إذا كان لديه اعتقاد بأن الوحدات التجريبية أكثر تجانساً داخل القطاعات منها بين القطاعات .

تمارين (٥)

(١) البيانات التالية لجدول تحليل التباين لتصميم تام العشوائية :

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات	٦	١٦,٩		
الخطأ				
الكلية	٤١	٤٥,٢		

المطلوب :

- أ — إكمال جدول تحليل التباين .
- ب — ما هو عدد المعالجات التي تشملها التجربة .
- ج — هل تؤيد هذه البيانات الفرض بوجود اختلاف بين متوسطات المعالجات . استخدم مستوى معنوية $\alpha = 0,01$.
- د — بفرض أن $\bar{m}_1 = 3,7$ ، $\bar{m}_2 = 4,1$ هل تدل هذه البيانات على وجود اختلاف بين متوسطي المجتمعين m_1 ، m_2 ؟ افترض أنه يوجد ٧ مشاهدات لكل معالجة ، استخدم مستوى معنوية $\alpha = 0,10$.
- هـ — أوجد فترة ثقة ٩٠٪ للفرق بين متوسطي المجتمعين $(m_2 - m_1)$.
- و — أوجد فترة ثقة ٩٠٪ لمتوسط المجتمع (m_1) .

(٢) أراد أحد المكاتب المتخصصة في أعمال مراجعة الدفاتر المالية للمؤسسات الكبرى تقييم الأتعاب التي يتقاضاها المكتب نظير الخدمات التي يقدمها، وكجزء من هذا التقييم أن يقارن تكاليف المراجعة للشركات ذات الأحجام المختلفة ، وقد قرر المكتب قياس حجم المؤسسة العملية بمقدار مبيعاتها السنوية، وبناء على ذلك تم تقسيم مجتمع العملاء إلى ثلاثة مجتمعات فرعية :

س = مجتمع العملاء الذين تزيد مبيعاتهم عن ٢٥٠ مليون جنيه .

ص = مجتمع العملاء الذين تتراوح مبيعاتهم بين ١٠٠ مليون ، ٢٥٠ مليون جنيه .

ع = مجتمع العملاء الذين تقل مبيعاتهم عن ١٠٠ مليون جنيه

وفد قام المكتب باختيار عينة عشوائية مكونة من عشرة عملاء من كل مجتمع من هذه المجتمعات ، والجدول التالي يلخص تكاليف المراجعة (بآلاف الجنيهات) :

تكاليف المراجعة (بآلاف الجنيهات)		
س	ص	ع
٢٥٠	١٠٠	٨٠
١٥٠	١٥٠	١٢٥
٢٧٥	٧٥	٢٠
١٠٠	٢٠٠	١٨٦
٤٧٥	٥٥	٥٢
٦٠٠	٨٠	٩٢
١٥٠	١١٠	٨٨
٨٠٠	١٦٠	١٤١
٣٢٥	١٣٢	٧٦
٢٣٠	٢٣٣	٢٠٠

المطلوب :

أ - استخدام تحليل التباين لتحديد ما إذا كان هناك فرق معنوي بين متوسطات تكاليف المراجعة للمجتمعات الثلاثة من العملاء ،
استخدم مستوى معنوية $\alpha = 0.05$

ب - إيجاد فترة ثقة ٩٥٪ للفرق (المين - المين) .

ج - أذكر الشروط اللازم توافرها حتى يمكن عمل الاستدلال الإحصائي في الجزئين أ ، ب .

(٣) أراد مدير أحد الشركات التي تقوم بالاستعانة بعدد كبير من مندوبي المبيعات دراسة تأثير نظام دفع الأجر على كمية المبيعات التي يحققها المندوب في الشهر. حيث يوجد ثلاثة أنظمة لدفع الأجر (العمولة ، الأجر الثابت ، أجر ثابت منخفض + عمولة) ولإجراء هذه الدراسة قام المدير بسحب عينة عشوائية من مندوبي المبيعات الذين يطبق عليهم الأنظمة المختلفة لدفع الأجر ، ويخلص الجدول التالي المبيعات (بالجنیهات) التي حققها هؤلاء المندوبين خلال الشهر :

عمولة	أجر ثابت	أجر ثابت + عمولة
٤٢٥	٤٢٠	٤٣٠
٥٠٧	٤٤٨	٤٩٢
٤٥٠	٤٣٧	٤٧٠
٤٩٣	٤٣٢	٥٠١
٤٦٦	٤٤٤	
٤٩٢		

المطلوب :

أ - هل تريد هذه البيانات وجود لاختلاف بين متوسط مبيعات المندوبين

في ظل الأنظمة المختلفة لدفع الأجر .

ب - أوجد فترة ثقة ٩٠٪ لمتوسط مبيعات المندوبين الذين يتقاضون

(أجر ثابت + عمولة) .

ج - إيجاد فترة ثقة ٩٠٪ للفرق في متوسط المبيعات للمندوبين الذين

يتقاضون (أجر ثابت + عمولة) وهؤلاء الذين يتقاضون أجر

ثابت .

(٤) الجدول التالي لتحليل التباين لتصميم قطاعات كاملة العشوائية :

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعاملات	٣	٢٨,٢		
القطاعات	٥		١٣,٨	
الخطأ		٣٤,١		
الكلي				

المطلوب :

أ - أكمل جدول تحليل التباين .

ب - هل تدل هذه البيانات على وجود لاختلاف بين متوسطات المعالجات .

ج - هل تريد هذه البيانات ضرورة استخدام تصميم القطاعات لهذه

التجربة .

د - إذا كانت الأوساط الحسابية للعينات ف ، ق هي $\bar{y}_F = 9,7$ ،

$\bar{y}_Q = 12,1$ على السترتيب ، أوجد فترة ثقة ٩٠٪ للفرق

($\mu_F - \mu_Q$) .

الفصل السادس

الاستدلال الإحصائي باستخدام أسلوب كا^٢

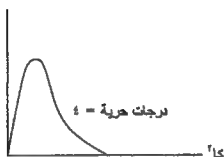
(٦ - ١) مقدمة :

استعرضنا في الفصول السابقة طرق عمل الاستدلال الإحصائي عن بعض معالم المجتمع بالاعتماد على توزيع z . لوت أوف . ولكن في بعض الأحيان قد لا نستطيع الاعتماد على مثل هذه التوزيعات . ونلجأ إلى ما يعرف بأسلوب كاي تربيع Chi-Square Technique . والذي يعتمد على توزيع كاي تربيع .

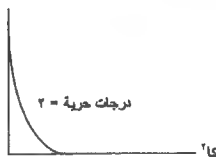
ويعتبر توزيع كاي تربيع (كا^٢) من التوزيعات المستمرة الهامة . ولإيضاح فكرة هذا التوزيع باختصار : نفترض أن لدينا متغيراً عشوائياً (س) له توزيع معتدل وسطه الحسابي يساوي μ وانحرافه المعياري يساوي σ . فإذا ما طرحنا من هذا المتغير وسطه الحسابي وقسمنا على انحرافه المعياري نحصل على متغير (z) يتبع للتوزيع المعتدل المعياري بوسط حسابي يساوي للصفر وانحراف معياري يساوي للوحد الصحيح . ويتربيع المتغير z نحصل على متغير يتبع لتوزيع مستمر شديد الالتواء تجاه اليمين وتتراوح قيم هذا المتغير بين صفر و ما لا نهاية . ويُعرف هذا التوزيع بتوزيع كا^٢ بدرجات حرية تساوي الواحد الصحيح . فإذا ما كان لدينا n من المتغيرات المستقلة z_1, z_2, \dots, z_n ، فإن مجموع مربعات هذه المتغيرات يتبع لتوزيع كا^٢ بدرجات حرية n . ومتوسط هذا التوزيع يساوي عدد درجات الحرية ، وببإينه يساوي ضعف عدد درجات الحرية . ويوضح للشكل (٦ - ١) توزيع كا^٢ عند

درجات حرية مختلفة . ويلاحظ أن التواء المنحنى تجاه اليمين يقل بزيادة عدد درجات الحرية ويقترب هذا للتوزيع من الاعتدال إذا كانت درجات الحرية أكبر من أو تساوي ٣٠ وتوجد جدول لتوزيع χ^2 تعطي قيم χ^2 المناظرة للمساحات المختلفة للطرف الأيمن ودرجات الحرية المختلفة .

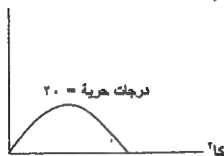
وسوف نتناول في هذا الفصل بعض التطبيقات على استخدام χ^2 وهي : الاستدلال الإحصائي عن أكثر من نمبتي مجتمعين ، اختبار الاستقلال بين متغيرين وصفيين ، اختبار جودة التوفيق ، الاستدلال الإحصائي عن تبالين المجتمع .



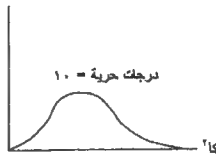
(ب)



(أ)



(د)



(ج)

شكل (٦ - ١)

٦١ - ٢) الاستدلال الإحصائي عن أكثر من نسبيتي مجتمعين :

إذا فرضنا أن أحد الشركات التجارية أرادت القيام بدراسة تفضيلات المستهلك لحملات زيوت الطعام المختلفة . وكان لدى هذه الشركة ثلاثة أنواع من الزيوت (أ ، ب ، ج) . فقامت باختيار عينة عشوائية مكونة من ١٥٠ مستهلكاً . وباستجواب كل مستهلك عن النوع الذي يفضلهُ ، أمكن الحصول على النتائج الموضحة بالجدول (٦ - ١) . هل تؤيد هذه البيانات تفضيل المستهلك لنوع معين ؟

جدول (٦ - ١)

نوع الزيت	أ	ب	ج	المجموع
عدد المستهلكين	٦١	٥٣	٣٦	١٥٠

للإجابة على هذا السؤال يجب التعرف أولاً على التوزيع الاحتمالي لهذه النوعية من البيانات. حيث أنها تتبع لتوزيع يُعرف بالتوزيع الاحتمالي المتعدد الحدود Multinomial وتتلخص خواصه في الآتي :

خواص التوزيع الاحتمالي المتعدد الحدود :
١ - عدد المحاولات التي تتكون منها للتجربة يساوي n
٢ - عدد النواتج الممكنة في كل محاولة يساوي k
٣ - احتمالات النواتج في كل محاولة ثابتة (أي لا تتغير من محاولة إلى أخرى) وتساوي $\theta_1, \theta_2, \dots, \theta_k$. حيث $\theta_1 + \theta_2 + \dots + \theta_k = 1$
٤ - المحاولات مستقلة .

ونلاحظ أن خواص للتوزيع المتعدد الحدود قريبة الشبه من خواص توزيع ذي الحدين . ونستطيع القول أن توزيع ذي الحدين هو حالة خاصة من التوزيع المتعدد الحدود (حيث $k = 2$) .

وغالبا ما تكون القيم الحقيقية للاحتتمالات $\theta_1, \theta_2, \dots, \theta_r$ مجهولة وبالتالي يكون هدفنا الأساسى هو عمل استدلال إحصائى عن هذه الاحتمالات . ففي مثالنا الحالى والخاص بتفضيل المستهلك والذي يحقق شروط التوزيع المتعدد الحدود إذا فرضنا أن :

- ١. θ_1 = نسبة المستهلكين الذين يفضلون النوع أ .
- ٢. θ_2 = نسبة المستهلكين الذين يفضلون النوع ب .
- ٣. θ_3 = نسبة المستهلكين الذين يفضلون النوع ج .

فيكون هدفنا هو اختبار فرض العدم بأنه لا يوجد تفضيل لأي نوع من الأنواع الثلاثة في مقابل البديل بأنه يوجد تفضيل لنوع أو أكثر من هذه الأنواع . وبالتالي يمكن صياغة هذه الفروض على الصورة الآتية :

$$H_0: \theta_1 = \theta_2 = \theta_3 = \frac{1}{3} \quad (\text{لا يوجد تفضيل}) .$$

$$H_1: \text{ يوجد نسبة واحدة على الأقل تزيد عن } \frac{1}{3} \quad (\text{يوجد تفضيل}) .$$

فإذا كان فرض العدم صحيحاً فإننا نتوقع أن يكون عدد المستهلكين لكل نوع يمثل $\frac{1}{3}$ حجم العينة تقريباً . بمعنى أنه إذا كانت n_1, n_2, n_3 ترمز على التوالي إلى عدد المستهلكين الذين يفضلون النوع أ ، النوع ب ، النوع جـ والتي سنطلق عليها التكرارات المشاهدة فإن التكرارات المتوقعة المناظرة لها يمكن إيجادها كالآتي :

$$\text{توقع } (n_1) = n \theta_1$$

$$= 150 \left(\frac{1}{3} \right) = 50 .$$

، كذلك أيضاً : توقع (٢ ن) = توقع (٢ ن) = ٥٠

وإحصائية الاختبار في هذه الحالة والتي تقيس درجة الاختلاف بين بيانات العينة (التكرارات المشاهدة) والبيانات تحت شرط صحة فرض العدم (التكرارات المتوقعة) هي :

$$\chi^2_{\text{كا}} = \frac{[I(٢ن) - \text{توقع}(٢ن)]^2}{\text{توقع}(٢ن)} + \frac{[I(١ن) - \text{توقع}(١ن)]^2}{\text{توقع}(١ن)} + \frac{[I(٥٠) - \text{توقع}(٥٠)]^2}{\text{توقع}(٥٠)}$$

ويلاحظ أنه كلما بُعُدت التكرارات المشاهدة عن المتوقعة ، كلما حصلنا على قيمة أكبر لإحصائية الاختبار $\chi^2_{\text{كا}}$. وهذا يعني أن القيمة الكبيرة لإحصائية الاختبار تعطي مؤشراً على عدم صحة فرض العدم .

وحتى نتمكن من أخذ القرار برفض أو قبول فرض العدم لابد أن تكون على علم بشكل التوزيع العيني لإحصائية الاختبار $\chi^2_{\text{كا}}$. وقد أمكن إثبات أنه تحت شرط صحة فرض العدم فإن التوزيع العيني لإحصائية الاختبار يقترب من توزيع $\chi^2_{\text{كا}}$ بدرجات حرية $v = 3 - 1 = 2$. وعلى فرض أننا نريد إجراء الاختبار عند مستوى معنوية $\alpha = 0.05$ ، فإننا سنرفض فرض العدم H_0 إذا كان :

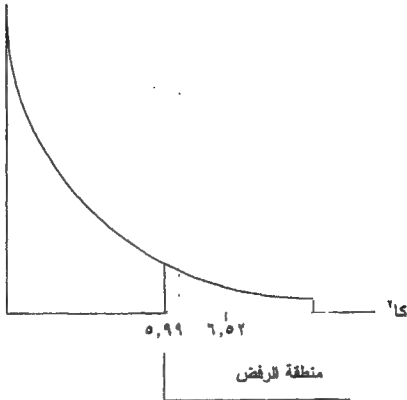
$$\chi^2_{\text{كا}} < \chi^2_{(٢, ٠.٠٥)} = ٥.٩٩ \quad (\text{من جدول كا}^2)$$

وباستخدام بيانات عينة المستهلكين التي تم سحبها عشوائياً نحصل على :

$$\chi^2_{\text{كا}} = \frac{[I(٥٠) - \text{توقع}(٥٠)]^2}{\text{توقع}(٥٠)} + \frac{[I(٥٠) - \text{توقع}(٥٠)]^2}{\text{توقع}(٥٠)} + \frac{[I(٥٠) - \text{توقع}(٥٠)]^2}{\text{توقع}(٥٠)}$$

$$٦.٥٢ = \frac{[٥٠ - ٣٦]^2}{٥٠} + \frac{[٥٠ - ٥٣]^2}{٥٠} + \frac{[٥٠ - ٦١]^2}{٥٠}$$

وحيث أن α^2 المحسوبة (٦,٥٢) تقع في منطقة الرفض (كما هو موضح بالشكل (٦ - ٢)) فإننا نستنتج رفض فرض العدم وقبول الفرض البديل بتفضيل المستهلك لنوع أو أكثر من أنواع لازيوت .



شكل (٦ - ٢)

ويمكن تلخيص عناصر اختبار الفرض الإحصائي المتعلق بأكثر من نسبي مجتمعين في الآتي :

اختبار الفرض الإحصائي المتعلق بأكثر من نسبيتي مجتمعين

الفروض الإحصائية :

$$H_0 : \theta_1 = \theta_2 , \theta_1 = \theta_3 , \theta_2 = \theta_3 , \dots , \theta_1 = \theta_k = \theta$$

(حيث $\theta_1 , \theta_2 , \dots , \theta_k$ هي القيم المفترضة لنسب المجتمعات)

H_1 : يوجد على الأقل نسبة واحدة لا تساوي القيمة المفترضة

إحصائية الاختبار : $\chi^2 = \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$

منطقة الرفض : $\chi^2 < (\chi^2_{\alpha, v})$ (حيث $v = k - 1$)

الشروط :

- ١ - أن يكون توزيع المجتمعات الأصلية معتدلاً .
 - ٢ - أن يكون حجم العينة كبيراً بدرجة تكفي لجعل التكرار المتوقع مساوياً للقيمة ٥ على الأقل .
- ويلاحظ أنه في حالة عدم توافر الشرط ٢ فأنتنا ندمج التكرارات المتوقعة وبالتالي التكرارات المشاهدة المناظرة لها وذلك حتى نستوفي ذلك الشرط . وممن الطبيعي أن يحسب عدد درجات الحرية بعد هذا الإنماج .

مثال (١) :

إذا كانت اللائحة الداخلية لإحدى الشركات تنص على نظام معين لتوزيع الحوافز السنوية على موظفي الشركة . حيث يعتمد هذا النظام على الدرجات التي تمنح للموظف بواسطة رئيسة المباحث في العمل . فالموظف الذي تزيد درجته عن ٨٠ يستحق الحد الأقصى للحوافز السنوية ، والذي تتراوح درجته بين ٥٠ ، ٨٠ يستحق الحوافز العادية ، والذي تقل درجته عن ٥٠ لا يستحق أي حوافز . وكانت الشركة قد وضعت في خطتها أن يستحق

الحد الأقصى ٢٥٪ من الموظفين ، يستحق الحوافز العادية ٦٥٪ . بينما لا تستحق النسبة الباقية (١٠٪) من الموظفين أي حوافز. ويعتمد سنة واحدة من استخدام هذا النظام تم سحب عينة عشوائية من ٦٠٠ موظف بالشركة . وكان توزيع الحوافز السنوية كما هو موضح بالجدول (٦ - ٢) . هل تدل هذه البيانات على وجود اختلاف معنوي بين توزيع الحوافز وفقا للاتحة الداخلية للشركة وتوزيعها وفقا لما وضعته الشركة في خطتها . استخدم مستوى معنوية $\alpha = 0.01$

جدول (٦ - ٢)

لا يستحق الحوافز	يستحق حوافز عادية	يستحق الحد الأقصى
٤٢	٣٦٥	١٩٣

الحل :

بفرض أن :

١٠ = نسبة الموظفين الذين لا يستحقون أي حوافز .

٢٠ = نسبة الموظفين الذين يستحقون حوافز عادية .

٣٠ = نسبة الموظفين الذين يستحقون الحد الأقصى للحوافز .

فبالتالي يكون فرض العدم والذي يمثل خطة الشركة في توزيع الحوافز

على الصورة :

$$H_0: 0.10 : 0.20 : 0.70 , \quad H_1: 0.20 : 0.30 : 0.50$$

ويكون للفرض البديل على الصورة :

H_1 : يوجد نسبة واحدة على الأقل لا تتفق مع خطة الشركة .

إحصائية الاختبار : χ^2 = مج (التكرار المشاهد - التكرار المتوقع)
التكرار المتوقع

منطقة الرفض : $\chi^2 < \chi^2_{(v, \alpha)} = \chi^2_{(v, 0.01)} = 9.21$

وباستخدام بيانات العينة يمكن حساب إحصائية الاختبار كما هو موضح بالجدول (٦ - ٣)

جدول (٦ - ٣)

التكرار المشاهد	التكرار المتوقع	مشاهد - متوقع	(مشاهد - متوقع) ^٢	(مشاهد - متوقع) ^٢ متوقع
٤٢	$60 = (0.10) 600$	١٨-	٣٢٤	٥,٤
٣٦٥	$390 = (0.65) 600$	٢٥-	٦٢٥	١,٦
١٩٣	$100 = (0.25) 600$	٤٣	١٨٤٩	١٢,٣
٦٠٠	٦٠٠	صفر		$19,3 = \chi^2$

وحيث أن قيمة χ^2 المحسوبة تقع في منطقة الرفض فأنا نستنتج رفض فرض العدم وقبول البديل بأن توزيع الحوافز السنوية لا يتفق مع خطة الشركة. وذلك عند مستوى معنوية $\alpha = 0.01$

وبالإضافة إلى إجراء الاختبار عن نسب المجتمعات فإنه يمكن إيجاد فترة ثقة لأي نسبة من هذه النسب . فعلى سبيل المثال يمكن إيجاد فترة ثقة ٩٥٪ لنسبة الموظفين الذين يستحقون الحد الأقصى للحوافز كالتالي :

$$\hat{p} \pm 1.96 \hat{\sigma}_{\hat{p}}$$

$$\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

$$\begin{aligned} \text{حيث } \hat{\theta} = \frac{\sum z}{n} = \frac{193}{600} = 0.32 \\ \sqrt{\frac{(0.32 - 1) \cdot 0.32}{600}} \sqrt{1.96 \pm 0.32} = \\ = 0.04 \pm 0.32 = \\ = (0.28, 0.36) \end{aligned}$$

وهذا يعني أن تتراوح نسبة الموظفين الذين يستحقون الحد الأقصى للحوافز فيما بين ٢٨٪ ، ٣٦٪ ومن ذلك يتضح أنه يجب زيادة عدد الدرجات اللازمة للحصول على الحد الأقصى للحوافز حتى يمكن تحقيق النسبة ٢٥٪ والتي وضعتها الشركة في خطتها .

(٦ - ٣) اختبار الاستقلال بين متغيرين وصفيين :

إذا فرضنا أن أحد الباحثين الاقتصاديين أراد دراسة العلاقة بين حجم السيارة المشتراة حديثاً والشركة المنتجة لها . فقام باختيار عينة عشوائية من ١٠٠٠ مشتري . وتم تصنيفهم وفقاً لحجم السيارة المشتراة والشركة المنتجة لها . ويوضح الجدول (٦ - ٤) البيانات التي تم الحصول عليها .

جدول (٦ - ٤)

الشركة المنتجة للسيارة	أ	ب	ج	د	المجموع
					حجم السيارة
صغير	١٥٧	٦٥	١٨١	١٠	٤١٣
متوسط	١٢٦	٨٢	١٤٢	٤٦	٣٩٦
كبير	٥٨	٤٥	٦٠	٢٨	١٩١
المجموع	٣٤١	١٩٢	٣٨٣	٨٤	١٠٠٠

والآن بفرض أن جدول (٦ - ٥) يمثل احتمال وقوع كل ناتج من نواتج الجدول (٦ - ٤) . حيث تشير $_{11}\theta$ إلى احتمال شراء سيارة صغيرة الحجم ومنتجة بواسطة الشركة أ ، $_{21}\theta$ تشير إلى احتمال شراء سيارة صغيرة الحجم ومنتجة بواسطة الشركة ب وهكذا بالنسبة لبقية خلايا الجدول . كما يشير مجموع الاحتمالات في كل صف أو كل عمود إلى ما يعرف بالاحتمالات الهامشية Marginal Probabilities . حيث $_{10}\theta$ تمثل احتمال شراء سيارة صغيرة الحجم ، $_{20}\theta$ تمثل احتمال شراء سيارة منتجة بواسطة الشركة أ . وهكذا بالنسبة لبقية الاحتمالات الهامشية .

جدول (٦ - ٥)

الشركة المنتجة للسيارة	أ	ب	جـ	د	المجموع	حجم السيارة
						صغير
	١١٠	٢١٠	٣١٠	٤١٠	١٠	
	١٢٠	٢٢٠	٣٢٠	٤٢٠	٢٠	متوسط
	١٣٠	٢٣٠	٣٣٠	٤٣٠	٣٠	كبير
	١٠	٢٠	٣٠	٤٠	١	المجموع

ولتحديد ما إذا كان هناك علاقة بين المتغيرين (حجم السيارة والشركة المنتجة لها) فإننا نرجع إلى تعريف الاستقلال بين المتغيرين في حالة الجدول المزدوج . حيث يقال أن المتغيرين مستقلان إذا كان الاحتمال في أي خلية يساوي حاصل ضرب الاحتمالات الهامشية المناظرة . وبناءً على ذلك فإننا نستطيع القول أنه إذا كان حجم السيارة مستقل عن الشركة المنتجة فسنجد أن :

$$١٠ \times ١٠ = ١١٠ \quad ، \quad ١٠ \times ١٠ = ١١٠$$

$$١٠ \times ١٠ = ١١٠ \quad ، \quad ١٠ \times ١٠ = ١١٠$$

وهكذا بالنسبة لبقية الاحتمالات في كل خلايا الجدول .

ولإجراء الاختبار الإحصائي لفرض الاستقلال بين المتغيرين نستخدم نفس الفكرة التي استخدمت في البند السابق (٦ - ٢) والتي تعتمد على إيجاد التكرار المتوقع لكل خلية بافتراض صحة هذا الفرض . حيث أن :

توقع (ن ١١) = التكرار المتوقع في الخلية التي تقع بالصف الأول والعمود الأول

$$= \text{ن } ١١$$

$$= \text{ن } ١٠ \times ١٠$$

$$\left\{ \begin{array}{l} \text{حيث } \text{ن } ١ = \text{مجموع الصف الأول} \\ \text{ن } ١ = \text{مجموع العمود الأول} \end{array} \right\} \quad \text{ن } = \left(\frac{\text{ن } ١}{\text{ن}} \right) \left(\frac{\text{ن } ١}{\text{ن}} \right)$$

$$= \frac{\text{ن } ١ \times \text{ن } ١}{\text{ن}}$$

وبنفس الأسلوب نحصل على :

$$\{ \text{حيث } \text{ن } ٢ = \text{مجموع العمود الثاني} \} \quad \text{توقع (ن ٢١)} = \frac{\text{ن } ١ \times \text{ن } ٢}{\text{ن}}$$

$$\{ \text{حيث } \text{ن } ٣ = \text{مجموع العمود الثالث} \} \quad \text{توقع (ن ٣١)} = \frac{\text{ن } ١ \times \text{ن } ٣}{\text{ن}}$$

⋮
⋮
⋮

$$\left\{ \begin{array}{l} \text{حيث } \text{ن } ١ = \text{مجموع الصف الثالث} \\ \text{ن } ٤ = \text{مجموع العمود الرابع} \end{array} \right\} \quad \text{توقع (ن ٤١)} = \frac{\text{ن } ٤ \times \text{ن } ١}{\text{ن}}$$

وباستخدام بيانات الجدول (٦ - ٤) نحصل على :

$$\text{توقع (ن ١١)} = \frac{٣٤١ \times ٤١٣}{١٠٠٠} = ١٤٠,٨٣٣$$

$$\text{توقع (ن ١٢)} = \frac{١٩٢ \times ٤١٣}{١٠٠٠} = ٧٩,٢٩٦$$

$$\text{توقع (ن ١٣)} = \frac{٣٨٣ \times ٤١٣}{١٠٠٠} = ١٥٨,١٧٩$$

⋮
⋮
⋮

$$\text{توقع (ن ١٤)} = \frac{٨٤ \times ١٩١}{١٠٠٠} = ١٦,٠٤٤$$

وبوضوح جدول (٦ - ٦) التكرارات المشاهدة والمتوقعة (بداخل

قوسين) أمثلنا الحالي :

جدول (٦ - ٦)

د	جـ	ب	أ	الشركة المنتجة للسيارة
				حجم السيارة
١٠ (٣٤,٦٩٢)	١٨١ (١٥٨,١٧٩)	٦٥ (٧٩,٢٩٦)	١٥٧ (١٤٠,٨٣٣)	صغير
٤٦ (٣٢,٢٦٤)	١٤٢ (١٥١,٦٦٨)	٨٢ (٧٦,٠٣٢)	١٢٦ (١٣٥,٠٣٦)	متوسط
٢٨ (١٦,٠٤٤)	٦٠ (٧٣,١٥٣)	٤٥ (٣٦,٦٧٢)	٥٨ (٦٥,١٣١)	كبير

وبعد إيجاد التكرارات المتوقعة نستخدم إحصائية χ^2 للمقارنة بين التكرار المشاهد والمتوقع في كل من خلايا الجدول كالآتي :

$$\chi^2 = \sum \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

$$\frac{(16,044 - 28)^2}{16,044} + \dots + \frac{(79,296 - 60)^2}{79,296} + \frac{(140,833 - 107)^2}{140,833} = 40,81$$

حيث تشير القيم الكبيرة لإحصائية χ^2 إلى وجود اختلاف كبير بين التكرارات المشاهدة والمتوقعة وهذا من شأنه التشكيك في صحة الفرض باستقلال المتغيرين . وللحكم على قيمة الإحصائية χ^2 فإنه يتم مقارنتها بقيمة χ^2 الجدولية بدرجات حرية مساوية (عدد الصفوف - 1) (عدد الأعمدة - 1) . ففي مثالنا الحالي تكون درجات الحرية = (3 - 1) (4 - 1) = 6 . وبالتالي فإنه عند مستوى معنوية $\alpha = 0,05$ ، نرفض الفرض باستقلال المتغيرين إذا كان :

$$\chi^2 < \chi^2_{(0,05)} = 12,59$$

وحيث أن قيمة χ^2 المحسوبة (40,81) أكبر من قيمة χ^2 الجدولية (12,59) فإننا نرفض فرض العلم باستقلال المتغيرين . وهذا يعني أنه توجد علاقة بين حجم السيارة المشتراة حديثا والشركة المنتجة لها .

ومما سبق نستطيع تلخيص اختبار الاستقلال بين متغيرين وصفين في الآتي :

اختبار الاستقلال بين متغيرين وصفيين

الفروض الإحصائية :

H_0 : لا توجد علاقة بين المتغيرين (المتغيران مستقلان)

H_1 : توجد علاقة بين المتغيرين (المتغيران غير مستقلين)

إحصائية الاختبار :

$$\chi^2 = \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

$$\chi^2 = \frac{\text{مجموع الصف} \times \text{مجموع العمود}}{\text{المجموع الكلي}}$$

منطقة الرفض : $\chi^2 > \chi^2_{(v, \alpha)}$

$$\text{حيث } v = (\text{عدد الصفوف} - 1) (\text{عدد الأعمدة} - 1)$$

الشرط : يجب أن يكون حجم العينة كبيراً بدرجة تكفي لجعل التكرار المتوقع في كل خلية لا يقل عن ٥ ، وفي حالة عدم توافر هذا الشرط فإننا نلجأ إلى إدماج التكرارات المتوقعة وبالتالي التكرارات المشاهدة المناظرة لها . وذلك حتى نستوفي هذا الشرط ونحسب درجات الحرية بعد هذا الإدماج .

مثال (٢) :

أراد قسم المراقبة على جودة الإنتاج في أحد الشركات أن يحدد ما إذا كان هناك علاقة بين سنوات الخبرة ومعدل الخطأ والذي يقاس بعدد الوحدات المعيبة في كل ١٠٠٠ وحدة منتجة بواسطة العامل . فقام باختيار عينة عشوائية مكونة من مائة عامل ويوضح جدول (٦ - ٧) نتائج هذه التجربة . فهل تؤيد هذه النتائج وجود علاقة بين سنوات الخبرة . ومعدل الخطأ في إنتاج العامل ؟
استخدم مستوى معنوية $\alpha = 0.05$.

جدول (٦ - ٧)

معدل الخطأ في الإنتاج	سنوت الخبرة	أقل من سنة	سنة إلى أقل من ٥	٥ - ١٠	المجموع
مرتفع	٦	٩	٩	٢٤	
متوسط	٩	١٩	٢٣	٥١	
منخفض	٧	٨	١٠	٢٥	
المجموع	٢٢	٣٦	٤٢	١٠٠	

الحل :

نبدأ أولاً بحساب التكرارات المتوقعة في كل خلية بافتراض صحة فرض العدم بأن المتغيرين مستقلان . أي أن :

$$\text{توقع (ن ١١)} = \frac{\text{مجموع الصف الأول} \times \text{مجموع العمود الأول}}{\text{المجموع الكلي}}$$

$$= \frac{٢٢ \times ٢٤}{١٠٠} = ٥,٢٨$$

$$\text{توقع (ن ٢١)} = \frac{\text{مجموع الصف الأول} \times \text{مجموع العمود الثاني}}{\text{المجموع الكلي}}$$

$$= \frac{٣٦ \times ٢٤}{١٠٠} = ٨,٦٤$$

وهكذا بالنسبة لجميع خلايا الجدول . يوضح جدول (٦ - ٨) التكرارات المشاهدة والمتوقعة (بدخل قوسين) .

جدول (٦ - ٨)

سنوات الخبرة / معدل الخطأ	أقل من سنة	سنة إلى أقل من ٥	٥ - ١٠
مرتفع	٦,٢٨ (٥,٢٨)	٩ (٨,٦٤)	٩ (١٠,٠٨)
متوسط	٩ (١١,٢٢)	١٩ (١٨,٣٦)	٢٣ (٢١,٤٢)
منخفض	٧ (٥,٥٠)	٨ (٩)	١٠ (١٠,٥٠)

والخطوة التالية لذلك هي إجراء اختبار الاستقلال بين المتغيرين :

الفروض الإحصائية :

H_0 : لا توجد علاقة بين سنوات الخبرة ومعدل الخطأ .

H_1 : توجد علاقة بين سنوات الخبرة ومعدل الخطأ .

إحصائية الاختبار :

$$\chi^2 = \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

منطقة الرفض :

$$\chi^2_{\alpha} = \chi^2_{(v, \alpha)} = \chi^2_{(4, 0.05)} = 9.49$$

$$\left[\begin{array}{l} \text{حيث } v = (\text{عدد الصفوف} - 1) (\text{عدد الأعمدة} - 1) \\ 4 = (1 - 3) (1 - 3) = \end{array} \right]$$

وباستخدام بيانات الجدول (٦ - ٨) يمكن حساب إحصائية الاختبار

χ^2 كما هو موضح بالجدول (٦ - ٩) .

جنول (٦ - ٩)

التكرار المشاهد	التكرار المتوقع	مشاهد - متوقع	(مشاهد - متوقع) ^٢	(مشاهد - متوقع) ^٣
٦	٥,٢٨	٠,٧٢	٠,٥١٨٤	٠,٠٩٨
٩	٨,٦٤	٠,٣٦	٠,١٢٩٦	٠,٠١٥
٩	١٠,٠٨	-١,٠٨	١,١٦٦٤	-١,١١٦
٩	١١,٢٢	-٢,٢٢	٤,٩٢٨٤	-٤,٤٣٩
١٩	١٨,٣٦	٠,٦٤	٠,٤٠٩٦	٠,٠٢٢
٢٣	٢١,٤٢	١,٥٨	٢,٤٩٦٤	٠,١١٧
٧	٥,٥٠	١,٥٠	٢,٢٥٠٠	٠,٤٠٩
٨	٩,٠٠	-١,٠٠	١,٠٠٠٠	-١,١١١
١٠	١٠,٥٠	٠,٥٠	٠,٢٥٠٠	٠,٠٢٤
١٠٠	١٠٠	صفر		١,٣٥١ = ك

وحيث أن قيمة ك^٢ المحسوبة (١,٣٥١) أقل من قيمة ك^٢ الجدولية (٩,٤٩) فإننا نستنتج أن بيانات العينة غير كافية لتأييد الفرض البديل بأنه توجد علاقة بين سنوات الخبرة ومعدل الخطأ في إنتاج العامل وذلك عند مستوى معنوية $\alpha = ٠,٠٥$.

(٦ - ٤) اختبار جودة التوفيق : Goodness of Fit test

يستخدم اختبار جودة التوفيق لتحديد التوزيع الاحتمالي الذي تتبع له بيانات المجتمع أو لاختبار مدى تبعية البيانات لتوزيع معين . ونورد فيما يلي بعض الأمثلة التطبيقية على استخدام هذا الاختبار .

(٦ - ٤ - ١) اختبار جودة التوفيق لتوزيع ذي الحدين :

مثال (٣) :

أرادت إحدى شركات الاستيراد والتصدير دراسة توزيع الوحدات للتالفة في شحنة من سلعة معينة وكانت معبأة في صناديق حيث يحتوي الصندوق الواحد على ٢ وحدات من السلعة... وقد أتى مصدر هذه السلعة أن الوحدات التالفة في هذه الشحنة تتبع لتوزيع ذي الحدين ولإختبار صحة هذا الادعاء تم فحص الوحدات التالفة في عينة عشوائية مكونة من ٢٠٠ صندوق ، وتم الحصول على النتائج الموضحة بالجدول (٦ - ١٠) المطلوب اختبار ادعاء مصدر هذه السلعة . استخدم مستوى معنوية ٠,٠٥ .

جدول (٦ - ١٠)

عدد الوحدات التالفة في الصندوق	صفر	١	٢	٣	٤	٥	٦	المجموع
عدد الصناديق	٣١	٥١	٧٠	٣٢	٩	٥	٢	٢٠٠

الحل :

حيث أننا نرغب في اختبار صحة إدعاء المصدر بأن الوحدات التالفة في الشحنة تتبع لتوزيع ذي الحدين فيجب أولاً تحديد معالم هذا التوزيع وهي عدد المحاولات (ن) ، واحتمال النجاح (θ) . بالنسبة إلى ن فنجد أنها تمثل أقصى رقم لعدد الوحدات التالفة في الصندوق وهو ٦. وبالنسبة لاحتمال النجاح (θ) فيمكن تقديره بإيجاد نسبة التالف في العينة . أي أن :

$$\hat{\theta} = \frac{\text{عدد للوحدات التالفة في العينة}}{\text{العدد الكلي للوحدات في العينة}}$$

$$\frac{\text{صفر} \times ٣١ + ٥١ \times ١ + ٧ \times ٢ + ٣٢ \times ٣ + ٩ \times ٤ + ٥ \times ٥ + ٢ \times ٦}{٦ \times ٢٠٠} = ٠,٣ =$$

ومن ذلك نستطيع إيجاد التكرارات المتوقعة بافتراض صحة فرض العدم بأن عدد الوحدات التالفة يتبع توزيع ذي الحدين كما هو موضح بجدول (٦ - ١١) .

جدول (٦ - ١١)

عدد الوحدات التالفة في الصندوق (س)	ح (س=س) = $\binom{٢٠٠}{س} (٠,٣)^س (٠,٧)^{٢٠٠-س}$ حيث ن=٦، ٠، ٣=٠,٣	التكرار المتوقع (ح × س = س)
صفر	ح (س= صفر) = $\binom{٢٠٠}{٠} (٠,٣)^٠ (٠,٧)^{٢٠٠} = ٠,١١٧٦$	$٢٣,٥٢ = ٠,١١٧٦ \times ٢٠٠$
١	ح (س= ١) = $\binom{٢٠٠}{١} (٠,٣)^١ (٠,٧)^{١٩٩} = ٠,٣٠٢٥$	$٦٠,٥٠ = ٠,٣٠٢٥ \times ٢٠٠$
٢	ح (س= ٢) = $\binom{٢٠٠}{٢} (٠,٣)^٢ (٠,٧)^{١٩٨} = ٠,٣٢٤١$	$٦٤,٨٢ = ٠,٣٢٤١ \times ٢٠٠$
٣	ح (س= ٣) = $\binom{٢٠٠}{٣} (٠,٣)^٣ (٠,٧)^{١٩٧} = ٠,١٨٥٢$	$٣٧,٠٤ = ٠,١٨٥٢ \times ٢٠٠$
٤	ح (س= ٤) = $\binom{٢٠٠}{٤} (٠,٣)^٤ (٠,٧)^{١٩٦} = ٠,٠٥٩٥$	$١١,٩٠ = ٠,٠٥٩٥ \times ٢٠٠$
٥	ح (س= ٥) = $\binom{٢٠٠}{٥} (٠,٣)^٥ (٠,٧)^{١٩٥} = ٠,٠١٠٢$	$٢,٠٤ = ٠,٠١٠٢ \times ٢٠٠$
٦	ح (س= ٦) = $\binom{٢٠٠}{٦} (٠,٣)^٦ (٠,٧)^{١٩٤} = ٠,٠٠٠٧$	$٠,١٤ = ٠,٠٠٠٧ \times ٢٠٠$

يلاحظ أننا دمجنا التكرارات المتوقعة عند س = ٤، ٥، ٦ حتى نستوفي شرط تطبيق كاي^٢ وهو ألا يقل التكرار المتوقع في أي خلية عن العدد ٥ .
والخطوة التالية لإيجاد التكرارات المتوقعة هي إجراء اختبار الفرض بأن البيانات تتبع لتوزيع ذي الحدين .
الفروض الإحصائية :

- H₀ : عدد الوحدات التالفة في الشحنة يتبع لتوزيع ذي الحدين .
- H₁ : عدد الوحدات التالفة لا يتبع لتوزيع ذي الحدين .

إحصائية الاختبار :

$$\chi^2 = \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

منطقة الرفض :

$$\chi^2_{\alpha} < \chi^2_{(v, \alpha)} = \chi^2_{(v, 0.05)} = 7.815$$

$$\left[\text{حيث } v = \text{عدد التكرارات المتوقعة} - 1 - \text{عدد المعالم التي تم تقديرها} \right]$$

$$3 = 1 - 1 - 0 =$$

وباستخدام بيانات الجدول (٦ - ١١) يمكن حساب إحصائية الاختبار

كما هو موضح بالجدول (٦ - ١٢) .

جدول (٦ - ١٢)

عدد الوحدات التألف في الصندوق	التكرار المشاهد	التكرار المتوقع	مشاهد - متوقع	مشاهد - متوقع) ^٢	مشاهد - متوقع) ^٢ متوقع
صفر	٣١	٢٣,٥٢	٧,٤٨	٥٥,٥٩٠	٢,٣٧٩
١	٥١	٦٠,٥٠	٩,٥٠-	٩٠,٢٥٠	١,٤٩٢
٢	٧٠	٦٤,٨٢	٥,١٨	٢٦,٨٣٢	٠,٤١٤
٣	٣٢	٣٧,٠٤	٥,٠٤-	٢٥,٤٠٢	٠,٦٨٦
٤ - ٦	١٦	١٤,٠٨	١,٩٢	٣,٦٨٦	٠,٢٦٢
	٢٠٠	٢٠٠	صفر		٥,٢٣٣ = χ^2

وحيث أن قيمة χ^2 المحسوبة (٥,٢٣٣) أقل من قيمة χ^2 الجدولية

(٧,٨١٥) فإننا نستنتج أن البيانات تتبع لتوزيع ذي الحدين .

(٦ - ٤ - ٢) اختبار جودة التوفيق لتوزيع بواسون :

مثال (٤) :

في إحدى الدراسات التي أجريت لمعرفة التوزيع الاحتمالي لعدد المرضى الذين يصلون خلال ساعة إلى العيادة الخارجية بأحد المستشفيات تم اختيار عينة عشوائية مكونة من ٥٠ ساعة عمل ، وتم تسجيل عدد المرضى الذين يصلون خلال الساعة . والجدول (٦ - ١٣) يلخص النتائج التي يمكن الحصول عليها . المطلوب اختبار فرض العدم بأن عدد المرضى الذين يصلون خلال الساعة يتبع توزيع بواسون بوسط حسابي $\lambda = ٣,٨$. استخدم معلوية ٠,٠١ .

جدول (٦ - ١٣)

عدد المرضى الذين يصلون خلال الساعة	صفر	١	٢	٣	٤	٥	٦	٧	٨ فأكثر	المجموع
عدد الساعات	صفر	١	٥	٨	١٥	٩	٧	٣	٢	٥٠

الحل :

نبدأ أولاً بإيجاد التكرارات المتوقعة بافتراض صحة فرض العدم بأن عدد المرضى الذين يصلون خلال الساعة يتبع لتوزيع بواسون بوسط حسابي $\lambda = ٣,٨$ (كما هو موضح بجدول (٦ - ١٤)) .

جدول (٦ - ١٤)

عدد المرضى الذين يصلون خلال الساعة (س)	ح (س = س) = $\frac{e^{-\lambda} \lambda^s}{s!}$ ، حيث $\lambda = ٣,٨$	التكرار المتوقع $e^{-\lambda} \lambda^s =$
صفر	ح (س = صفر) = $\frac{e^{-٣,٨} ٣,٨^0}{0!} = ٠,٠٢٢٤$	$٠,٠٢٢ \times ٠,٠٢٢٤ \times ٥٠ = ٠,١٢$
١	ح (س = ١) = $\frac{e^{-٣,٨} ٣,٨^1}{1!} = ٠,٠٨٥٠$	$٤,٢٥ \times ٠,٠٨٥٠ \times ٥٠ = ٩,٢٥$
٢	ح (س = ٢) = $\frac{e^{-٣,٨} ٣,٨^2}{2!} = ٠,١٦١٥$	$٨,٠٧٥ \times ٠,١٦١٥ \times ٥٠ = ٦٦,١٥$
٣	ح (س = ٣) = $\frac{e^{-٣,٨} ٣,٨^3}{3!} = ٠,٢٠٤٦$	$١٠,٢٣ \times ٠,٢٠٤٦ \times ٥٠ = ١٠٤,٦٣$
٤	ح (س = ٤) = $\frac{e^{-٣,٨} ٣,٨^4}{4!} = ٠,١٩٤٤$	$٩,٧٢ \times ٠,١٩٤٤ \times ٥٠ = ٩٤,٧٢$
٥	ح (س = ٥) = $\frac{e^{-٣,٨} ٣,٨^5}{5!} = ٠,١٤٧٧$	$٧,٣٨٥ \times ٠,١٤٧٧ \times ٥٠ = ٥٤,٧٧$
٦	ح (س = ٦) = $\frac{e^{-٣,٨} ٣,٨^6}{6!} = ٠,٠٩٣٦$	$٤,٦٨ \times ٠,٠٩٣٦ \times ٥٠ = ٩٣,٦٨$
٧	ح (س = ٧) = $\frac{e^{-٣,٨} ٣,٨^7}{7!} = ٠,٠٥٠٨$	$٢,٥٤ \times ٠,٠٥٠٨ \times ٥٠ = ٥٠,٨٠$
٨ فأكثر	ح (س ≤ ٨) = ١ - ح (س > ٨) $١ - ٠,٠٤٠٠ = ٠,٩٦$	$٢,٠٠ \times ٠,٠٤٠٠ \times ٥٠ = ٤٠,٠٠$

والخطوة التالية اختيار الفرض بأن البيانات تتبع لتوزيع بواسون بوسط

$$\text{حسابي } \lambda = 3,8 :$$

الفروض الإحصائية :

H_0 : عدد المرضى الذين يصلون خلال الساعة يتبع لتوزيع بواسون

$$\text{بوسط حسابي } = 3,8$$

H_1 : عدد المرضى الذين يصلون خلال الساعة لا يتبع لتوزيع بواسون

$$\text{بوسط حسابي } = 3,8$$

إحصائية الاختبار :

$$\chi^2 = \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

منطقة الرفض :

$$\chi^2_{\alpha} < \chi^2_{(v, \alpha)} = \chi^2_{(v, 0,01)} = 10,576$$

$$\{ \text{حيث } v = 6 - 1 = 5 \}$$

وباستخدام بيانات الجدول (6 - 14) يمكن حساب إحصائية الاختبار

χ^2 كما هو موضح بالجدول (6 - 15) .

جول (٦ - ١٥)

عدد المرضى الذين يصلون خلال ساعة	التكرار المشاهد	التكرار المتوقع	مشاهد - متوقع	٢ (مشاهد - متوقع)	١ (مشاهد - متوقع)
صفر - ١	١	٥,٣٧	٤,٣٧-	١٩,٠٩٧	٣,٥٥٦
٢	٥	٨,٠٧٥	٣,٠٧٥-	٩,٤٥٦	١,١٧١
٣	٨	١٠,٢٣	٢,٢٣٠-	٤,٩٧٣	٠,٤٨٦
٤	١٥	٩,٧٢	٥,٢٨	٢٧,٨٧٨	٢,٨٦٨
٥	٩	٧,٣٨٥	١,٦١٥	٢,٦٠٨	٠,٣٥٣
٦ فأكثر	١٢	٩,٢٢	٢,٧٨	٧,٧٢٨	٠,٨٣٨
المجموع	٥٠	٥٠	صفر		٩,٢٧٢ = كا

وحيث أن قيمة كا^٢ المحسوبة (٩,٢٧٢) أقل من قيمة كا^٢ الجدولية (١٥,٠٨٦) فإننا نستنتج أن عدد المرضى الذين يصلون خلال الساعة إلى العيادة الخارجية بالمستشفى يتبع لتوزيع بواسون بوسط حسابي $\lambda = ٣,٨$.

(٦ - ٤ - ٣) اختبار جودة التوفيق للتوزيع المعتدل :

مثال (٥) :

أراد أحد الباحثين الاقتصاديين دراسة للتوزيع الاحتمالي للدخل السنوي لطائفة المحامين . فقام باختيار عينة عشوائية مكونة من ٩٠ محام ؛ وخلص جدول (٦ - ١٦) النتائج التي تم الحصول عليها . المطلوب اختبار الفرض بأن الدخل السنوي لطائفة المحامين يتبع لتوزيع معتدل بوسط حسابي $\mu = ٥٠$ ألف جنيه ، انحراف معياري $\sigma = ١٠$ آلاف جنيه . استخدم مستوى معنوية ٠,٠٥ .

جدول (٦ - ١٦)

فئات الدخل السنوي (بالآلاف الجنيهات)	أقل من ١٠	-١٠	-٢٠	-٣٠	-٤٠	-٥٠	-٦٠	-٧٠	٨٠ فأكثر	المجموع
عدد المحامين	٥	٩	١٥	٢٣	٢٠	٨	٦	٣	١	٩٠

الحصل :

نبدأ أولاً بإيجاد التكرارات المتوقعة بافتراض صحة فرض العدم بأن الدخل السنوي يتبع لتوزيع معتدل بوسط حسابي $\mu = ٥٠$ ألف جنيهها ، انحراف معياري $\sigma = ١٠$ آلاف جنيهها (كما هو موضح بجدول (٦ - ١٧)) .

جدول (٦ - ١٧)

فئات الدخل السنوي (بالآلاف الجنيهات)	الاحتمال (المساحة تحت المنحنى المعتدل المعياري) حيث $Z = \frac{\mu - \text{م.س.}}{\sigma} = \frac{٥٠ - \text{م.س.}}{١٠}$	التكرار المتوقع $= ٩٠ \times \text{ح(س)}$
أقل من ١٠	$\text{ح(س} > ١٠) = \text{ح(س} > ٤) = ٠,٠٠١٤$	$٩٠ \times \text{صفر} = \text{صفر}$
-١٠	$\text{ح(١٠} > \text{س} > ٢٠) = \text{ح(٤} > Z > ٣) = ٠,٠٠١٤$	$٠,٠٠١٤ \times ٩٠ = ٠,١٢٦$
-٢٠	$\text{ح(٢٠} > \text{س} > ٣٠) = \text{ح(٣} > Z > ٢) = ٠,٠٢١٤$	$٠,٠٢١٤ \times ٩٠ = ١,٩٢٦$
-٣٠	$\text{ح(٣٠} > \text{س} > ٤٠) = \text{ح(٢} > Z > ١) = ٠,١٣٥٩$	$٠,١٣٥٩ \times ٩٠ = ١٢,٢٣١$
-٤٠	$\text{ح(٤٠} > \text{س} > ٥٠) = \text{ح(١} > Z > ٠) = ٠,٣٤١٣$	$٠,٣٤١٣ \times ٩٠ = ٣٠,٧١٧$
-٥٠	$\text{ح(٥٠} > \text{س} > ٦٠) = \text{ح(صفر} > Z > ١) = ٠,٣٤١٣$	$٠,٣٤١٣ \times ٩٠ = ٣٠,٧١٧$
-٦٠	$\text{ح(٦٠} > \text{س} > ٧٠) = \text{ح(٧٠} > Z > ١) = ٠,١٣٥٩$	$٠,١٣٥٩ \times ٩٠ = ١٢,٢٣١$
-٧٠	$\text{ح(٧٠} > \text{س} > ٨٠) = \text{ح(٣} > Z > ٢) = ٠,٠٢١٤$	$٠,٠٢١٤ \times ٩٠ = ١,٩٢٦$
٨٠ فأكثر	$\text{ح(س} > ٨٠) = \text{ح(٣} < Z) = ٠,٠٠١٤$	$٠,٠٠١٤ \times ٩٠ = ٠,١٢٦$

والخطوة التالية هي اختبار الفرض بتبعية للبيانات للتوزيع المعتدل
بوسط حسابي $\mu = 50$ ألف جنيه ، انحراف معياري $\sigma = 10$ آلاف جنيه :

الفروض الإحصائية :

H_0 : الدخل السنوي لطائفة المحامين يتبع لتوزيع معتدل بوسط حسابي

$\mu = 50$ ألف جنيه ، انحراف معياري $\sigma = 10$ آلاف جنيه

H_1 : الدخل السنوي لا يتبع لتوزيع معتدل بوسط حسابي $\mu = 50$ ألف ..

جنيها ، انحراف معياري $\sigma = 10$ آلاف جنيه

إحصائية الاختبار :

$$\chi^2 = \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

منطقة الرفض :

$$\chi^2_{\alpha} < \chi^2_{(v, \alpha)} = \chi^2_{(2, 0.05)} = 5.991$$

$$\{ \text{حيث } v = 4 - 1 = 3 \}$$

وباستخدام بيانات الجدول (٦ - ١٧) يمكن حساب إحصائية الاختبار

χ^2 كما هو موضح بالجدول (٦ - ١٨) .

جدول (٦ - ١٨)

فئات الدخل	التكرار المشاهد	التكرار المتوقع	مشاهد - متوقع	(مشاهد - متوقع) ^٢	(مشاهد - متوقع) ^٢
أقل من ٤٠	٥٢	١٤,٢٨٣	٣٧,٧١٧	١٤٢٢,٥٧٢	٩٩,٥٩٩
٤٠ -	٢٠	٣٠,٧١٧	١٠,٧١٧-	١١٤,٨٥٤	٣,٧٣٩
٥٠ -	٨	٣٠,٧١٧	٢٢,٧١٧-	٥١٦,٠٦٢	١٦,٨٠١
٦٠ فأكثر	١٠	١٤,٢٨٣	٤,٢٨٣-	١٨,٣٤٤	١,٢٨٤
المجموع	٩٠	٩٠	صفر		١٢١,٤٢٣ = كآ

وحيث أن قيمة كآ^٢ للمحسوبة (١٢١,٤٢٣) أكبر من قيمة كآ^٢ الجدولية (٧,٨١٥) فإننا نرفض فرض العدم . وهذا لا يعني بالضرورة عدم تبعية الدخل السنوي لطائفة المحامين للتوزيع المعتدل بل من الممكن أن يتبع الدخل السنوي للتوزيع المعتدل ولكن بوسط حسابي يختلف عن ٥٠ ألف جنيهها وانحراف معياري يختلف عن ١٠ آلاف جنيهها .

(٦ - ٤ - ٤) اختبار جودة التوفيق للتوزيع المنتظم :

مثال (٦) :

قام أحد المسؤولين عن مراكز المطافئ بأحد المدن بتسجيل آخر ٧٠٠ حريق نشب في المدينة وكان توزيعهم على مدار أيام الأسبوع كما هو موضح بالجدول (٦ - ١٩) . المطلوب اختبار فرض العدم بأن حدوث الحرائق في المدينة يتوزع توزيعاً منتظماً على مدار أيام الأسبوع . استخدم مستوى معنوية ٠,٠١ .

جدول (٦ - ١٩)

أيام الأسبوع	المبت	الأحد	الاثنين	الثلاثاء	الأربعاء	الخميس	الجمعة	المجموع
عدد الحرائق	٨٥	١٢٩	٧٢	٩١	١٢٣	١٣٥	٦٥	٧٠٠

الحل :

الفروض الإحصائية :

H_0 : تتوزع الحرائق توزيعاً منتظماً على مدار أيام الأسبوع .

H_1 : لا تتوزع الحرائق توزيعاً منتظماً على مدار أيام الأسبوع .

$$\chi^2 = \frac{\text{التكرار الملاحظ} - \text{التكرار المتوقع}}{\text{التكرار المتوقع}}$$

$$\text{منطقة الرفض : } \chi^2 < \chi^2_{(7,0.05)} = \chi^2_{(10,0.01)} = 16,812$$

ولحساب إحصائية الاختبار يجب أولاً حساب التكرار المتوقع بافتراض صحة فرض العدم بأن عدد الحرائق يتوزع بالتساوي على أيام الأسبوع كما هو موضح بجدول (٦ - ٢٠) .

جدول (٦ - ٢٠)

أيام الأسبوع	التكرار الملاحظ	التكرار المتوقع	ملاحظ - متوقع	(ملاحظ - متوقع) ^٢	(ملاحظ - متوقع) ^٢ / متوقع
السبت	٨٥	١٠٠	-١٥	٢٢٥	٢,٢٥
الأحد	١٢٩	١٠٠	٢٩	٨٤١	٨,٤١
الاثنين	٧٢	١٠٠	-٢٨	٧٨٤	٧,٨٤
الثلاثاء	٩١	١٠٠	-٩	٨١	٠,٨١
الأربعاء	١٢٣	١٠٠	٢٣	٥٢٩	٥,٢٩
الخميس	١٣٥	١٠٠	٣٥	١٢٢٥	١٢,٢٥
الجمعة	٦٥	١٠٠	-٣٥	١٢٢٥	١٢,٢٥
المجموع	٧٠٠	٧٠٠	صفر		$\chi^2 = 49,10$

وحيث أن قيمة كا^٢ المحسوبة (٤٩,١٠) أكبر من قيمة كا^٢ الجدولية (١٦,٨١٢) فإننا نرفض فرض العدم ونقبل البديل بأن الحرائق لا تتوزع بانتظام على مدار أيام الأسبوع وذلك عند مستوى معنوية ٠,٠١ .

وبالإضافة إلى ما سبق من أمثلة على اختبار جودة التوفيق لبعض التوزيعات المعروفة فإنه من الممكن استخدام هذا الاختبار أيضا لأي توزيع آخر لا يتبع إلى مثل هذه التوزيعات المعروفة . والمثال التالي يوضح ذلك .

مثال (٧) :

أرادت إدارة برامج التلفزيون مقارنة أنماط عادات المشاهدة لبرامج التلفزيون في عام ١٩٨٠ بما كانت عليه عام ١٩٧٠ م . فقامت باختيار عينة عشوائية من ١٠٠ مشاهد في عام ١٩٧٠ ، عام ١٩٨٠ وجدول (٦ - ٢١) يلخص النتائج التي تم الحصول عليها ، فهل تريد هذه البيانات اختلاف أنماط عادات المشاهدة في عام ١٩٨٠ عنها في عام ١٩٧٠ ، استخدم مستوى معنوية ٠,٠٥ .

جدول (٦ - ٢١)

عدد المشاهدتين		عدد ساعات مشاهدة التلفزيون في الأسبوع
١٩٨٠	١٩٧٠	
٩	٣	صفر -
٣	٤	- ٣
٩	١٠	- ٥
٢١	٢٠	- ٧
٤٤	٤٥	- ١٥
١٤	١٨	٢٥ فأكثر
١٠٠	١٠٠	المجموع

الحل :

الفروض الإحصائية :

H_0 : لا تختلف أنماط عادات المشاهدة في عام ١٩٨٠ عنها في عام ١٩٧٠ .

H_1 : تختلف أنماط عادات المشاهدة في عام ١٩٨٠ عنها في عام ١٩٧٠ .

$$\chi^2 = \frac{(\text{التكرار للمشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

وبافتراض صحة فرض العدم تكون التكرارات المتوقعة هي عدد المشاهدين في عام ١٩٧٠ .

منطقة الرفض :

$$\chi^2_{\alpha} < \chi^2_{(v, \alpha)} = \chi^2_{(4, 0.05)} = 9.488$$

{ حيث $v = 5 - 1 = 4$ لأننا نمجنا الفئتين الأولى والثانية حتى لا يقل التكرار المتوقع عن ٥ } .

وباستخدام بيانات العينة يمكن حساب إحصائية الاختبار χ^2 كما هو موضح بجدول (٦ - ٢٢) .

جدول (٦ - ٢٢)

عدد ساعات مشاهدة التلفزيون في الأسبوع	التكرار المشاهد	التكرار المتوقع	مشاهد - متوقع	(مشاهد - متوقع) ^٢	(مشاهد - متوقع) ^٣ متوقع
صفر -	١٢	٧	٥	٢٥	٣,٥٧١
- ٥	٩	١٠	- ١	١	٠,١٠٠
- ٧	٢١	٢٠	١	١	٠,٠٥٠
- ١٥	٤٤	٤٥	- ١	١	٠,٠٢٢
٢٥ فأكثر	١٤	١٨	- ٤	١٦	٠,٨٨٩
المجموع	١٠٠	١٠٠	صفر		٤,٦٣٢ = Σ

وحيث أن Σ^2 المحسوبة (٤,٦٣٢) أقل من Σ الجدولية
(٩,٤٨٨) فإننا نستنتج أن أنماط عادات المشاهدة في عام ١٩٨٠ لم تختلف
عنها في عام ١٩٧٠ .

(٤ - ٥) الاستدلال الإحصائي عن تباين المجتمع σ^2 :

من المعروف أن $\Sigma^2 = \text{مج} (س - \bar{س})^2 / (ن - ١)$ يستخدم كمقدر غير
متحيز لتباين المجتمع σ^2 . ولكن لكي نحدد درجة ثقة في التقدير لا بد من
معرفة شكل التوزيع العيني للمقدر Σ^2 . وقد تم إثبات أن للكمية :

$$\frac{\Sigma^2 (ن - ١)}{\sigma^2} \text{ تتبع لتوزيع } \Sigma^2 \text{ بدرجات حرية } (ن - ١)$$

ومن ذلك يمكن ائتمنان فترة ثقة ١٠٠ (١ - α) % لتباين المجتمع كالاتي :

$$\begin{aligned}
\alpha - 1 &= (1 - 0, \frac{\sigma}{\sigma})^2 \text{كا} > \text{كا} > (1 - 0, \frac{\sigma}{\sigma-1})^2 \text{كا} \\
\alpha - 1 &= (1 - 0, \frac{\sigma}{\sigma})^2 \text{كا} > \frac{(1-n)^2 \sigma^2}{\sigma^2} > (1 - 0, \frac{\sigma}{\sigma-1})^2 \text{كا} \\
\alpha - 1 &= (\frac{1}{(1 - 0, \frac{\sigma}{\sigma-1})^2 \text{كا}} > \frac{\sigma^2}{(1-n)^2 \sigma^2} > \frac{1}{(1 - 0, \frac{\sigma}{\sigma})^2 \text{كا}}) \\
\alpha - 1 &= (\frac{(1-n)^2 \sigma^2}{(1 - 0, \frac{\sigma}{\sigma-1})^2 \text{كا}} > \sigma^2 > \frac{(1-n)^2 \sigma^2}{(1 - 0, \frac{\sigma}{\sigma})^2 \text{كا}})
\end{aligned}$$

<p>فترة ثقة ١٠٠ ($\alpha - 1$) % لتباين المجتمع σ^2</p> $\frac{(1-n)^2 \sigma^2}{(1 - 0, \frac{\sigma}{\sigma-1})^2 \text{كا}} > \sigma^2 > \frac{(1-n)^2 \sigma^2}{(1 - 0, \frac{\sigma}{\sigma})^2 \text{كا}}$

مثال (٨) :

أراد مدير أحد شركات صناعة الأنوية معرفة تباين وزن قرص الدواء الجديد . فقام باختيار عينة عشوائية مكونة من ٣٠ قرص . وحسب تباين وزن القرص فكان ٣ ميلليجرام . فعلى فرض أن أوزان هذه الأقراص تتبع لتوزيع معتدل تقريبا فأوجد فترة ثقة ٩٠ % لتباين وزن قرص الدواء .

الحل :

فترة ثقة ١٠٠ ($\alpha - 1$) % لتباين المجتمع σ^2 هي :

$$\frac{(1-n)^2 \sigma^2}{(1 - 0, \frac{\sigma}{\sigma-1})^2 \text{كا}} > \sigma^2 > \frac{(1-n)^2 \sigma^2}{(1 - 0, \frac{\sigma}{\sigma})^2 \text{كا}}$$

وحيث أن درجات الحرية = $n - 1 = 30 - 1 = 29$ ، $\alpha = 0,10$ ،

فيستخدم جداول كا^2 نحصل على :

$$\text{كا}^2 (1 - 0.05) = \text{كا}^2 (0.95) = 42,007$$

$$\text{كا}^2 (1 - 0.05) = \text{كا}^2 (0.95) = 17,708$$

وبالتالي فإن فترة ثقة ٩٠٪ لتباين وزن قرص الدواء تكون على

الصورة :

$$\frac{(29)^3}{17,708} > \sigma^2 > \frac{(29)^3}{42,007}$$

$$4,91 > \sigma^2 > 2,04$$

وهذا يعني أننا نتق بدرجة ٩٠٪ في أن يقع تباين وزن قرص الدواء

بين ٢,٠٤ ، ٤,٩١ ميللجرام .

وبالإضافة إلى إيجاد فترة ثقة لتباين المجتمع فإنه يمكن إجراء

اختبارات الفروض الإحصائية عن تباين المجتمع .

اختبار الفرض الإحصائي عن تباين المجتمع (σ^2)	
اختبار ذو طرفين	اختبار ذو طرف واحد
الفروض الإحصائية :	الفروض الإحصائية :
$\sigma^2 = \sigma_0^2 : H_0$	$\sigma^2 = \sigma_0^2 : H_0$
$\sigma^2 \neq \sigma_0^2 : H_1$	$\sigma^2 > \sigma_0^2 : H_1$
	(أو $\sigma^2 < \sigma_0^2 : H_1$)
	حيث σ_0^2 القيمة المفترضة لتباين المجتمع
إحصائية الاختبار : $\chi^2 = \frac{(1-n)}{\sigma_0^2}$	إحصائية الاختبار : $\chi^2 = \frac{(1-n)}{\sigma_0^2}$
منطقة الرفض : $\text{كا}^2 > \text{كا}^2 (1 - 0.05)$	منطقة الرفض : $\text{كا}^2 > \text{كا}^2 (1 - 0.05)$
أو $\text{كا}^2 < \text{كا}^2 (1 - 0.05)$	أو $\text{كا}^2 < \text{كا}^2 (1 - 0.05)$
	إذا كان $\sigma^2 < \sigma_0^2 : H_1$
الشرط : أن يكون توزيع المجتمع المسحوب منه العينة معتدلاً (أو قريباً من الاعتدال) .	

مثال (٩) :

أراد مدير أحد البنوك أن يطبق سياسة الصف الواحد في تقديم الخدمة للعملاء بحيث يدخل العميل في الصف بمجرد وصوله للبنك وبعد ذلك يتم توزيع العميل على الشباك المختص بتقديم الخدمة . وقد وجد أنه بالرغم من أن هذه السياسة لن تؤثر على متوسط وقت انتظار العميل في البنك إلا أن المدير يفضل هذه السياسة لأنها تقلل من تباين وقت الانتظار . وكان يتوقع أن يكون هذا التباين أقل من تباين وقت الانتظار في حالة استخدام سياسة الصفوف المتعددة (حيث كان التباين = ٦٤) وللتأكد من ذلك قام بتطبيق سياسة الصف الواحد على عينة عشوائية مكونة من ٣٠ عميل وحسب الانحراف المعياري لوقت انتظار العميل فكان مساويا ٣ دقائق . فهل تؤيد بيانات العينة اعتقاد المدير . استخدم مستوى معنوية ٠,٠٥ .

الحل :

حيث أننا نرغب في تحديد ما إذا كان تباين وقت انتظار العميل في البنك يقل عن ٦٤ دقيقة فإن نوع الاختبار الذي يجب استخدامه هو اختبار ذو طرف أيسر وتكون عناصره كالآتي :

الفروض الإحصائية :

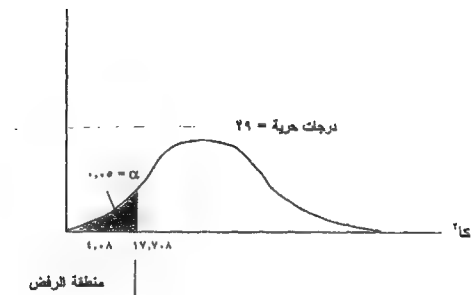
$$H_0 : \sigma^2 = 64$$

$$H_1 : \sigma^2 < 64$$

إحصائية الاختبار :

$$K = \frac{s^2 (n-1)}{\sigma^2}$$

منطقة الرفض $\chi^2_{\alpha} > \chi^2_{(1-\alpha, n-1)} = \chi^2_{(1-0.05, 29)} = 17.108$
 (كما هو موضح بالشكل (٦ - ٣))



شكل (٦ - ٣)

وباستخدام بيانات العينة نحصل على :

$$\chi^2_{\text{ع}} = \frac{(1 - n) \sigma^2}{\sigma^2}$$

$$4.08 = \frac{(29)(9)}{64} =$$

وحيث أن قيمة $\chi^2_{\text{ع}}$ المحسوبة (٤,٠٨) تقع في منطقة الرفض (كما هو واضح في الشكل (٦ - ٣)) فإننا نقبل للفرض البديل بأن $\sigma^2 > 64$ وبالتالي يكون اعتقاد المدير صحيحا في أن سياسة الصف للواحد سوف تؤدي إلى تقليل تباين وقت انتظار العميل .

مثال (١٠) :

أعلنت إحدى شركات التشييد والمباني أن تباين قوة ضغط الخرسانة المسلحة للمباني التي تقوم بإنشائها يزيد عن ٨٠ كيلو جرام لكل متر مربع وتم اختبار عينة عشوائية مكونة من ٢٥ عجيبة للخرسانة وحسب الانحراف المعياري لقوة الضغط فكان ٨ كيلو جرام لكل متر مربع . فهل تؤيد هذه البيانات إدعاء الشركة . استخدم مستوى معنوية ٠,٠٥ .

الحل :

الفروض الإحصائية :

$$H_0: \sigma^2 = 80$$

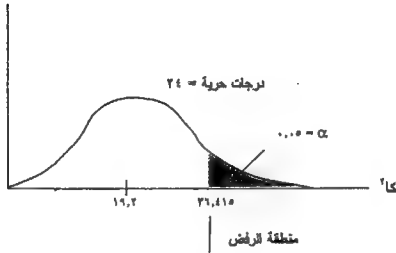
$$H_1: \sigma^2 < 80$$

إحصائية الاختبار :

$$\chi^2_{\alpha} = \frac{(n-1)s^2}{\sigma_0^2}$$

$$\text{منطقة الرفض } \chi^2_{\alpha} < \chi^2_{(1-\alpha)} = \chi^2_{(1-0.05)} = \chi^2_{(0.95)} = 36,415$$

(كما هو موضح بالشكل (٦ - ٤))



شكل (٦ - ٤)

وباستخدام بيانات العينة نحصل على :

$$كا^2 = \frac{ع^2 (ن - ١)}{٥\sigma^2} = \frac{(٦٤) (٢٤)}{٨٠} = ١٩,٢$$

وحيث أن قيمة كا^٢ المحسوبة (١٩,٢) تقع في منطقة القبول فإننا نستنتج أن بيانات العينة غير كافية لتأييد الفرض البديل وذلك عند مستوى معنوية ٠,٠٥ .

مثال (١١) :

يعتبر صافي حساب النقدية اليومية للشركة من الأمور التي تسترعي اهتمام مجلس الإدارة . وكان رئيس مجلس إدارة إحدى الشركات يرى ضرورة دراسة تباين صافي حساب النقدية اليومية للشركة والذي أثبتت الدراسات السابقة أنه يساوي ١٠٠ ألف جنيه . في حين أنه كان يعتقد أن هذا التباين يختلف بدرجة كبيرة عن هذه القيمة . وللتحقق من ذلك قام باختيار عينة عشوائية مكونة من صافي حساب النقدية لليومية لفترة حديثة شملت ١٤ يوما . وكانت النتائج كالموضحة بجدول (٦ - ٢٣) هل تؤيد هذه البيانات اعتقاد رئيس مجلس الإدارة . استخدم مستوى معنوية ٠,١٠ .

جنول (٦ - ٢٣)

اليوم	صافي حساب النقدية اليومية (بالآلاف الجنيهات)
١	١٧-
٢	٢٢+
٣	٥+
٤	صفر
٥	٨+
٦	٣-
٧	١٣+
٨	٢٠+
٩	٢٥+
١٠	٤١-
١١	صفر
١٢	٤-
١٣	٨+
١٤	١٣+

الحل :

حيث أننا نهدف لمعرفة ما إذا كان تباين صافي حساب النقدية اليومية للشركة يختلف عن ١٠٠ ألف جنيه فإن نوع الاختبار الذي يجب استخدامه هو اختبار ذو طرفين وتكون عناصره كالآتي :

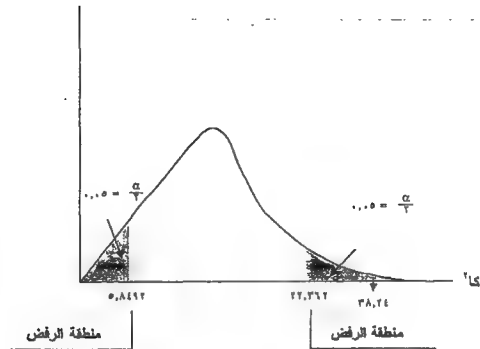
$$100 \neq \sigma^2 : H_1 \quad 100 = \sigma^2 : H_0$$

إحصائية الاختبار :

$$\frac{\sum_{i=1}^n (1 - \sigma_i)^2}{\sigma^2} = \chi^2_{\alpha}$$

منطقة الرفض $\chi^2_{\alpha} > (1 - \sigma_i)^2$ أو $\chi^2_{\alpha} < (1 - \sigma_i)^2$

$\chi^2_{\alpha} > 5.892 = (13, \dots, 90)$ أو $\chi^2_{\alpha} < 22.362 = (13, \dots, 90)$



شكل (٦ - ٥)

ولحساب إحصائية الاختبار يجب أولاً حساب تباين العينة σ^2

(كما هو موضح في جدول (٦ - ٢٤)) .

جول (٦ - ٢٤)

س	س - س	^٢ (س - س)
١٧-	٢٠,٥-	٤٢٠,٢٥
٢٢+	١٨,٥	٣٤٢,٢٥
٥+	١,٥	٢,٢٥
صفر	٣,٥-	١٢,٢٥
٨+	٤,٥	٢٠,٢٥
٣-	٦,٥-	٤٢,٢٥
١٣+	٩,٥	٩٠,٢٥
٢٠+	١٦,٥	٢٧٢,٢٥
٢٥+	٢١,٥	٤٦٢,٢٥
٤١-	٤٤,٥-	١٩٨٠,٢٥
صفر	٣,٥-	١٢,٢٥
٤-	٧,٥-	٥٦,٢٥
٨+	٤,٥	٢٠,٢٥
١٣+	٩,٥	٩٠,٢٥
٤٩	صفر	٣٨٢٣,٥٠

$$\bar{s} = \frac{\text{مجموع } s}{n} = \frac{٤٩}{١٤} = ٣,٥$$

$$s^2 = \frac{\sum (s - \bar{s})^2}{n - 1} = \frac{٣٨٢٣,٥}{١٣} = ٢٩٤,١٢$$

$$\therefore \text{كا} = \frac{s^2}{100} = \frac{(١٣) ٢٩٤,١٢}{100} = ٣٨,٢٤$$

وحيث أن كاً^٢ المحسوبة تقع في منطقة للرفض (كما هو موضح بالشكل (٦ - ٥)) فإننا نرفض فرض العدم . وبالتالي تؤيد البيانات اعتقاد رئيس مجلس الإدارة باختلاف تباين صافي حساب النقدية اليومية للشركة عن ١٠٠ ألف جنيه . ولحتمال أن يكون هذا القرار خاطئاً ٠,١٠ .

تمارين (٦)

(١) أرادت إدارة إحدى الشركات التجارية معرفة الوسيلة الأكثر فاعلية في الإعلان عن الأوكازيون السنوي الذي تقوم به . فقامت باختيار عينة عشوائية من زبائن الأوكازيون وباستجوابهم عن الوسيلة التي أعلنوا بها عن الأوكازيون تم الحصول على النتائج الموضحة بالجدول التالي :

الوسيلة	التلفزيون	الراديو	الصحف والمجلات	الاتصال الشخصي بين الأفراد
عدد الزبائن	٥٣	٣٢	٣٦	٤٨

المطلوب :

أ - هل تؤيد هذه البيانات وجود اختلاف بين نسب الزبائن الذين أعلنوا بالوسائل الأربعة ؟ استخدم $\alpha = ٠,١٥$

ب - أوجد فترة ثقة ٩٠ ٪ لنسبة الزبائن الذين أعلنوا بالأوكازيون عن طريق الاتصال الشخصي بين الأفراد .

(٢) لمعرفة تأثير أحد الأدوية الجديدة في علاج مرض معين تم اختيار عينة عشوائية من ٢٠٠ مريض تناولوا هذا الدواء فوجد من بينهم ٩٠ مريض تحسنت حالتهم الصحية ، ٤٠ مريض لم تتغير حالتهم الصحية ، ٣٠ مريض حدث لهم آثار جانبية ثانوية ، ٤٠ مريض حدث لهم آثار جانبية رئيسية . فهل تتفق هذه البيانات مع توقع الشركة المنتجة للدواء بأن ٥٠ ٪ من المستخدمين للدواء سوف تتحسن حالتهم ، ٣٠ ٪ لن تتغير حالتهم الصحية ، ١٥ ٪ تحدث لهم آثار جانبية ثانوية ، ٥ ٪ تحدث لهم آثار جانبية رئيسية ، استخدم $\alpha = ٠,٠١$.

(٣) أرادت منظمة العمل دراسة مشكلة العمال في ثلاث صناعات رئيسية حدث فيها إحلال الآلات محل العمل اليدوي . فقامت باختيار عينة عشوائية من مائة عامل من كل صناعة من للصناعات والذين فقدوا عملهم بسبب التقدم للتكنولوجي وإحلال الآلات محل العمل اليدوي وتم استجواب كل عامل عما إذا كان قد وجد عمل آخر دخل نفس الشركة أو في شركة جديدة وفي نفس الصناعة أو في صناعة جديدة أو لم يجد أي عمل آخر ، والجدول التالي يلخص النتائج التي تم الحصول عليها .

الوضع الحالي للعمال في الصناعة	نفس الشركة	شركة جديدة (نفس الصناعة)	صناعة جديدة	لم يجد عمل
أ	٦٢	١١	٢٠	٧
ب	٤٥	٨	٣٨	٩
جـ	٦٨	١٩	٨	٥

هل تؤيد هذه البيانات وجود علاقة بين الوضع الحالي للعامل والصناعة التي كان يعمل بها وتركها نتيجة للتقدم للتكنولوجي وإحلال الآلات محل العمل اليدوي ؟ استخدم مستوى معنوية $\alpha = 0.01$.

٤ التالي يبين توزيع مائة شخص أصيبوا بأزمة قلبية حسب الجنس

الجنس	نكر	أقضى	المجموع
أقل من ٣٠ سنة	٦	٤	١٠
٣٠ - ٦٠	٣٨	٤٢	٨٠
أكثر من ٦٠	٦	٤	١٠
المجموع	٥٠	٥٠	١٠٠

هل تؤيد هذه البيانات وجود علاقة بين نوع الشخص المصاب بأزمة قلبية وعمره . استخدم مستوى معنوية $\alpha = 0.10$.

(٥) قامت إحدى شركات الصناعات الدوائية بتحضير نوع جديد من الأدوية لعلاج الأرق ، وأرادت أن تقدر تباين الوقت الذي يمر بين تناول المريض للدواء ، وشعوره بالراحة والاسترخاء . وتم تجربة هذا الدواء على ٢٠ مريض . فوجد أن الانحراف المعياري للوقت الذي يمر بين تناول المريض للدواء وشعوره بالراحة والاسترخاء يساوي ١١ دقيقة . أوجد فترة ثقة ٩٠ % لتباين الوقت الذي يمر بين تناول المريض للدواء وشعوره بالراحة والاسترخاء .

(٦) يستخدم أحد منتجي الغلال آلة معينة لتعبئة منتج على أساس أن يكون الانحراف المعياري لوزن العبوة يساوي ١٠ جرام . ولكن لاحظ المنتج أن الانحراف المعياري لوزن بعض العبوات يختلف عن ١٠ جرام . فقام باختيار عينة عشوائية من ١٥ عبوة وحسب الانحراف المعياري فكان مساويا ١٥ جرام .

المطلوب :

أ — إيجاد فترة ثقة ٩٥ % لتباين وزن العبوة .

ب — هل تؤيد هذه البيانات رأي المنتج بأن الانحراف المعياري لوزن العبوة يختلف عن ١٠ جرام . استخدم مستوى معنوية ١٠ % .

(٧) تعاهد أحد مراكز البحوث على استيراد شحنة من الفخثران وقد أعلن المصدر أن تباين وزن الفخثر على الأكثر ٤ جرام . فقام أحد الباحثين باختيار عينة عشوائية من ٧ فخران ووجد أن الانحراف المعياري لوزن

الفأر ٦ جرام . فهل تؤيد هذه البيانات إدعاء المصدر للفئران . استخدم مستوى معنوية $\alpha = 0,05$.

(٨) أراد أحد الطيارين دراسة توزيع عدد محركات الطائرة التي تحتاج إلى بنزين إضافي بعد ٧ ساعات طيران . فقام باختيار عينة عشوائية مكونة من ٢٠٠ رحلة طيران وسجل عدد المحركات التي تحتاج إلى بنزين إضافي بعد ٧ ساعات طيران في كل رحلة ، والجدول التالي يلخص النتائج التي تم الحصول عليها :

عدد المحركات التي تحتاج إلى بنزين بعد ٧ ساعات طيران	صفر	١	٢	٣	٤	المجموع
عدد رحلات الطيران	١٥	٤٤	٧٥	٥٦	١٠	٢٠٠

المطلوب : اختبار الفرض بأن عدد محركات الطائرة التي تحتاج إلى بنزين بعد ٧ ساعات طيران يتبع لتوزيع ذي الحدين باحتمال نجاح $\theta = 0,05$. استخدم مستوى معنوية $\alpha = 0,01$.

(٩) أراد مراقب جودة الإنتاج أن يختبر فرض العدم بأن عدد الوحدات المعيبة في صندوق يحتوي على ٣ وحدات يتبع لتوزيع ذي الحدين باحتمال نجاح $\theta = 0,2$ ، فقام باختيار عينة عشوائية من ٢٠٠ صندوق والجدول التالي يلخص للنتائج التي تم الحصول عليها :

عدد الوحدات المعيبة في الصندوق	صفر	١	٢	٣	المجموع
عدد الصناديق	٤٦	٦٠	٥٨	٣٦	٢٠٠

المطلوب : إجراء الاختبار عند مستوى معنوية $\alpha = 0,10$.

(١٠) أرادت هيئة النقل والمواصلات دراسة توزيع عدد الحوادث التي يرتكبها سائق الأتوبيس سنويا فقامت باختيار عينة عشوائية مكونة من ١١٢٩٥٦ سائق وتم تسجيل عدد الحوادث التي ارتكبها للسائق في السنة وكانت كالمبينة بالجدول التالي :

عدد الحوادث التي ارتكبها السائق في السنة	صفر	١	٢	٣	المجموع
عدد السائقين	١٠٣٦٢٨	٧٣٨٩	١٩١٠	٢٩	١١٢٩٥٦

المطلوب : اختبار فرض العدم بأن عدد الحوادث التي يرتكبها السائق في السنة يتبع لتوزيع بواسون . استخدم مستوى معنوية $\alpha = ٠,٠٥$.

(١١) الجدول التالي يبين توزيع عدد الأخطاء في الساعة لثلاثين ساعة طيران :

عدد الأخطاء في الساعة	عدد الساعات
صفر	٣
١	٨
٢	٥
٣	٧
٤	٢
٥	١
٦	٢
٧	١
٨	صفر
٩	صفر
١٠	١
المجموع	٣٠

المطلوب : اختبار فرض العدم بأن هذه البيانات تتبع لتوزيع بواسون . استخدم مستوى معنوية $\alpha = ٠,٠١$.

(١٢) في إحدى الدراسات للتسويقية لختيرت عينة عشوائية مكونة من ٥٠٠ مستهلك لأنواع مختلفة من الشاي وتم استجواب كل منهم عن النوع الأكثر تفضيلا لديه والجدول التالي يلخص نتائج الدراسة :

النوع الأكثر تفضيلا	عدد المستهلكين
أ	٩١
ب	١٠٩
جـ	٨٥
د	١٠٠
هـ	١١٥
المجموع	٥٠٠

اختبر فرض العدم بأن تفضيلات المستهلك للأنواع المختلفة من الشاي تتبع للتوزيع المنتظم . استخدم مستوى معنوية $\alpha = ٠,١٠$.

(١٣) ألقيت زهرة نرد ١٢٠ مرة وكان عدد مرات ظهور الست أوجه من ١ إلى ٦ على الترتيب ٥ ، ٣٦ ، ٨ ، ٩ ، ٧ ، ٥٥ . المطلوب اختبار فرض العدم بأن الست أوجه في كل الرميات الممكنة لزهرة النرد تتوزع بانتظام . استخدم مستوى معنوية $\alpha = ٠,٠٥$.

(١٤) أراد أحد الباحثين الاقتصاديين دراسة التوزيع الاحتمالي للدخل الشهري لمهندسي المباني فقام باختيار عينة عشوائية من ١٠٠٠ مهندس وتم تسجيل الدخل الشهري لكل منهم ، والجدول التالي يلخص للنتائج التي تم الحصول عليها :

فئات الدخل الشهري	أقل من ٨٠٠	٨٠٠ -	٨٥٠ -	٩٠٠ -	٩٥٠ -	١٠٠٠ فأكثر	المجموع
عدد المهندسين	٢٦	١٤٦	٣٦١	٣١١	١٤٣	١٣	١٠٠٠

اختبر فرض العدم بأن الدخل الشهري لمهندسي المباني يتبع لتوزيع
معتدل بوسط حسابي $\mu = 900$ وانحراف معياري $\sigma = 50$.

الفصل السابع

الارتباط الخطي بين الظواهر

١-٧ مقدمة:

تعتبر مقاييس الارتباط والتي نقدمها في هذا الفصل من الأدوات الهامة التي تجيب على العديد من الأسئلة المتعلقة بطبيعة العلاقات بين المتغيرات، وسوف نقتصر مناقشتنا في هذا الفصل على قياس قوة العلاقة الخطية بين متغيرين، وبالتالي سوف نقوم باستخدام تعبير "الارتباط الخطي بين المتغيرات"،

٢-٧ معاملات الارتباط وخصائصها

يتم تحليل الارتباط الخطي دائماً على أساس حساب ما يسمى بمعامل الارتباط والذي نرمز له بالرمز "r"، ويتصف معامل الارتباط بأن قيمته المطلقة لا تتجاوز الواحد الصحيح،

$$|r| \leq 1$$

أو بمعنى آخر

$$-1 \leq r \leq 1$$

ونستخلص من قيمة معامل الارتباط ما يلي:

- 1- إذا كانت قيمة معامل الارتباط تساوي صفر أو قريبة من الصفر فإننا نستنتج عدم وجود علاقة خطية بين المتغيرين، ويلاحظ هنا أننا ننفي

وجود العلاقة الخطية لأنه قد توجد حالات نجد فيها أن $r = 0$ صفر بينما توجد علاقة غير خطية بين المتغيرين كما سيتضح من الأمثلة فيما بعد.

2- إذا كانت إشارة معامل الارتباط موجبة فإن هذا يعني وجود علاقة خطية طردية بين المتغيرين، وإذا كانت إشارته سالبة دل ذلك على وجود علاقة عكسية بين المتغيرين.

3- إذا كانت $r = 1$ أو -1 دل ذلك على وجود علاقة خطية تامة بين المتغيرين وهي الحالة التي نجد فيها أن جميع للنقاط تقع على استقامة واحدة كما أسلفنا الذكر في الفصل السابق.

4- كلما اقتربت قيمة معامل الارتباط من الواحد الصحيح كلما زادت قوة العلاقة بين المتغيرين وكلما بعدت عن الواحد الصحيح واقتربت من الصفر ضعفت العلاقة بين المتغيرين، وبصورة تقريبية يمكن القول بأن العلاقة تعتبر قوية إذا زالت القيمة العددية لمعامل الارتباط عن $0,8$ وتعتبر العلاقة متوسطة إذا انحصرت القيمة العددية لمعامل الارتباط بين $0,5$ و $0,8$ وتكون ضعيفة إذا هلت عن $0,5$.

ونتناول فيما يلي بعض مقاييس معاملات الارتباط والتي تناسب الأنواع المختلفة للمتغيرات الإحصائية.

٧-٣ معامل بيرسون للارتباط

يستخدم معامل بيرسون لقياس الارتباط الخطي بين المتغيرات الكمية ولا يصلح للاستخدام في حالة البيانات النوعية، وتكون إحدى الصيغ الممكنة لحساب معامل بيرسون للارتباط الخطي بين متغيرين على الصورة

$$r = 1 - \frac{1}{\frac{1}{(1-n)^2} \text{مجم } (Z_{ص} - Z_{ص})^2} \quad (1-7)$$

حيث تعبر $Z_{ص}$ و $Z_{ص}$ عن الوحدات المعيارية المناظرة لمشاهدات كل من المتغيرين $ص$ و $ص$ على الترتيب وتعرف على الصورة:

$$Z_{ص} = \frac{(ص - ع_{ص})}{ع_{ص}} \quad (2-7)$$

$$Z_{ص} = \frac{(ص - ع_{ص})}{ع_{ص}} \quad (3-7)$$

وعند حساب الوحدات المعيارية لقيم أي متغير، نجد أن وسطها الحسابي دائماً ما يساوي الصفر وتباينها يساوي الواحد الصحيح، وبالتالي فهي تستخدم لمقارنة قيم الظواهر المختلفة على أساس "معيارى" موحد يستبعد تأثير الفروق بين المستوى العام لقيم المتغيرات (بجعل متوسطاتها الحسابية تساوي الصفر)، ويستبعد كذلك تأثير اختلاف درجات تباين قيم كل متغير (بجعل انحرافات المعيارية تساوي الواحد الصحيح).

ويلاحظ القارئ هنا أننا استخدمنا أدلة سفلية لكل من الوحدات المعيارية والانحرافات المعيارية وذلك لبيان ما إذا كان المقياس المشار إليه يخص المتغير $ص$ أم المتغير $ص$.

ويمكن إعادة كتابة المعادلة (1-7) على الصورة

$$r = 1 - \frac{1}{\frac{1}{(1-n)^2} \left[\text{مجم } Z_{ص}^2 + \text{مجم } Z_{ص} Z_{ص} + \text{مجم } Z_{ص}^2 \right]}$$

وحيث أن

$$1 = \frac{ع_{ص}^2}{ع_{ص}^2} = \frac{مج_{ص} (س - س_{ص})^2}{ع_{ص}^2 (1 - ن)^2} = \frac{مج_{ص} Z_{ص}^2}{1 - ن}$$

وبنفس المنطق نجد أن

$$\frac{مج_{ص} Z_{ص}^2}{1 - ن}$$

وبالتالي تكون

$$r = 1 - \frac{1}{2} \left[\frac{مج_{ص} Z_{ص}^2}{1 - ن} - 2 \right]$$

(٧-٤)

$$\frac{مج_{ص} Z_{ص}^2}{1 - ن} = r$$

و بالتعويض عن $Z_{ص}$ و $س$ من المعادلتين (٧-٢) و (٧-٣) في المعادلة

(٧-٤) يمكن كتابة معامل بيرسون للارتباط على الصورة

$$r = \frac{\frac{1}{(n-1)} \text{مج} (س - \bar{س})(ص - \bar{ص})}{ع س ع ص} \quad (٥-٧)$$

ويطلق على البسط في الصيغة السابقة اسم التغاير والذي نرسم له بالرمز ع س ع ص ، بالتالي يمكن كتابة معامل بيرسون للارتباط على الصورة

$$r = \frac{ع س ص}{ع س ع ص} \quad (٦-٧)$$

ويمكن اشتقاق صورة بديلة لمعامل بيرسون للارتباط وذلك بإعادة كتابة صيغة التغاير كما يلي

$$r = \frac{\text{مج} (س - \bar{س})(ص - \bar{ص})}{ع س ع ص}$$

ويمكن إثبات أن

$$r = \frac{1}{n-1} \left(\text{مج} س ص - \frac{(\text{مج} س)(\text{مج} ص)}{n} \right) \quad (٧-٧)$$

ولتسهيل إجراء العمليات الحسابية في حالة ما إذا كانت قيم أحد المتغيرين أو كلاهما كبيرة، نناقش الآن تأثير إضافة مقدار ثابت وكذلك تأثير الضرب في مقدار ثابت على القيم المعيارية وبالتالي على قيمة معامل الارتباط.

افترض أننا قمنا بإضافة مقدار ثابت أ على كل قيمة من قيم س ثم ضربنا الناتج في مقدار ثابت ل بحيث نحصل على وحدات متغير جديد وليكن ي والذي يأخذ الصورة

$$ي = ل (س + أ)$$

وتكون مشاهدات المتغير الجديد ي على الصورة

$$ي_1 = ل (س_1 + أ)$$

$$ي_2 = ل (س_2 + أ)$$

$$ي_3 = ل (س_3 + أ)$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$ي_n = ل (س_n + أ)$$

والآن إذا أردنا حساب الوسط الحسابي لقيم المتغير ي فإننا نقوم بأخذ مجموع المعادلات السابقة، ونلاحظ هنا أنه عند أخذ مجموع الطرف الأيسر من المعادلات أن المقدار ل يكون عاملاً مشتركاً وبالتالي نحصل على الصيغة

$$م_ج ي = ل م_ج (س + أ)$$

$$م_ج ي = ل (م_ج س + م_ج أ)$$

وحيث أن أ مقدار ثابت وأن مجموع المقدار الثابت يساوي الثابت مضروباً في عدد الحدود، فإننا نحصل على الصيغة

$$م_ج ي = ل (م_ج س + ن أ)$$

بالتالي لحساب الوسط الحسابي لقيم ي نقسم الطرفين على ن لنحصل على العلاقة

$$\frac{\text{مـ ي}}{\text{ن}} = \frac{\text{مـ س}}{\left(\frac{\text{ن}}{\text{ن}} + \frac{\text{ن}}{\text{ن}}\right)}$$

$$\text{ي} = \text{ل} (\text{س} + \text{أ})$$

وتعني هذه النتيجة أنه للحصول على الوسط الحسابي لقيم ي نقوم بإجراء نفس التحويلات السابقة التي أجريناها على قيم س ونطبقها على الوسط الحسابي س، أي نقوم بإضافة أ ثم الضرب في ل،

مثال ٧-١

إذا كان الوسط الحسابي لقيم المتغير س هو $\bar{س} = 23$ ، أوجد الوسط الحسابي للمتغير ي والذي يعرف على الصورة

$$\text{ي} = ٠,٢ (\text{س} + ٧)$$

الحل:

من العلاقة السابقة يكون الوسط الحسابي للمتغير ي على الصورة

$$\bar{\text{ي}} = ٠,٢ (\bar{\text{س}} + ٧)$$

$$\bar{\text{ي}} = ٠,٢ (٢٣ + ٧)$$

$$\bar{\text{ي}} = ٦.$$

يلاحظ في هذا المثال أن قيمة الوسط الحسابي للمتغير الجديد ي تتأثر بقيمتي الثابتين أ و ل.

وإذا أردنا دراسة العلاقة بين تباين قيم ي وتباين قيم س، فإننا نكتب أولاً صيغة تباين قيم ي ثم نقوم بالتعويض بدلالة قيم س،

$$ع^٢ = \frac{١}{١-ن} \text{ مجـ } (ي - \overline{ي})^٢$$

$$ع^٢ = \frac{١}{١-ن} \text{ مجـ } (ل(س+١) - ل(\overline{س}+١))^٢$$

$$ع^٢ = \frac{١}{١-ن} \text{ مجـ } (ل س - ل \overline{س})^٢$$

$$ع^٢ = ل^٢ = \frac{١}{١-ن} \text{ مجـ } (س - \overline{س})^٢$$

$$ع^٢ = ل^٢ = ع^٢$$

وبالتالي يكون الانحراف المعياري لقيم ي على الصورة

$$ع = ل$$

أي أن الانحراف المعياري للمتغير الجديد ي لا يتأثر بعملية الجمع (لا يعتمد على الثابت الجمعي أ) ولكنه يتأثر بعملية الضرب (يعتمد على ل).
مثال ٧-٢

أوجد قيمة الانحراف المعياري للمتغير ي في المثال السابق إذا علمت أن الانحراف المعياري للمتغير س هو ١٠ .
الحل:

حيث أن

$$ي = ٢ + (س + ٧)$$

فإن الانحراف المعياري لقيم ي يكون

$$ع_ي = ٠,٢ ع_س$$

$$١٠ \times ٠,٢ =$$

$$٢ =$$

لاحظ هنا أننا تجاهلنا ثابت الجمع، ٧، تماماً.

والآن نقوم بدراسة تأثير عمليتي جمع وضرب مقدار ثابت، على قيم متغير، على الوحدات المعيارية له وذلك بإيجاد العلاقة بين $Z_س$ و $Z_ي$ ،

$$Z_ي = \frac{ع_ي - \bar{ع_ي}}{ع_ي}$$

$$= \frac{ل(س + ١) - ل(س + ١)}{ل ع_س}$$

$$= \frac{ل س - ل \bar{س}}{ل ع_س}$$

$$Z_س =$$

وتعني النتيجة السابقة أن القيم المعيارية لمتغير معين لا تتأثر بعمليات جمع مقدار ثابت على قيم متغير معين أو ضربها في مقدار ثابت، وحيث أن صيغة معامل بيرسون للارتباط في المعادلة (٧-١) تعتمد فقط على القيم المعيارية للمتغيرين فإننا نستنتج أن قيمة معامل الارتباط هي الأخرى لن تتأثر بعمليات الجمع والضرب، بالتالي إذا وجدنا أن قيم أحد المتغيرين أو كلاهما كبيرة مما يجعل العمليات الحسابية معقدة فيمكن أن نقوم بطرح مقدار ثابت من قيم كل من المتغيرين، وإذا احتوت القيم على عامل مشترك فيمكن أيضاً القسمة عليه ولن تختلف قيمة معامل الارتباط عند حسابها من القيم الأصلية أو القيم الجديدة.

لعل القارئ يتساءل عن سبب تعدد صيغ حساب معامل بيرسون للارتباط، يرجع السبب في ذلك إلى أن هذه الصيغ تخدم أغراضاً مختلفة، فالصيغة المعطاة في المعادلة (١_٧) تستخدم في تفسير معنى هذا المعامل في ظل الدرجات المختلفة للعلاقات الخطية بين المتغيرين، أما الصيغة (١_٧) فتكون مناسبة عند تحليل العلاقة بين الارتباط والانحدار الذي نقوم بدراسته في الفصل القادم، وحيث أن حساب الوحدات المعيارية لقيم المتغيرين ينطوي على عمليات حسابية مطولة، فإننا عادة ما نستخدم أحد الصيغتين (٥_٧) أو (٨_٧) لحساب قيمة معامل بيرسون للارتباط، وننصح باستخدام الصيغة (٥_٧) إذا وجدنا أن قيمتي الوسط الحسابي للمتغيرين S و $ص$ عددين صحيحين ونقوم باستخدام الصيغة (٨_٧) فيما عدا ذلك.

نبدأ فيما يلي بعرض بعض الأمثلة لإرساء المفاهيم السابقة وبيان الخطوات المتبعة لحساب قيمة معامل بيرسون للارتباط، وسوف نقوم في الأمثلة الثلاثة الأولى بتطبيق الصيغة (١_٧) لإبراز بعض الحالات الهامة.

مثال ٣-٧

احسب معامل بيرسون للارتباط باستخدام المشاهدات التالية عن متغيرين S و $ص$ ،

6	7	4	7	10	2	S
15	17	11	17	23	7	$ص$

الحل:

لحساب القيم المعيارية نبدأ أولاً بحساب الوسط الحسابي والانحراف المعياري لكل من قيم S و $ص$ ،

مسلسل	س	ص	س-٦	ص-١٥	(س-٦)²	(ص-١٥)²
١	٢	٧	٤-	٨-	١٦	٦٤
٢	١٠	٢٣	٤	٨	١٦	٦٤
٣	٧	١٧	١	٢	١	٤
٤	٤	١١	٢-	٤-	٤	١٦
٥	٧	١٧	١	٢	١	٤
٦	٦	١٥	٠	٠	٠	٠
المجموع	٣٦	٦٠	صفر	صفر	٣٦	١٥٦

$$\overline{س} = ٦ \div ٣٦ = ٦$$

$$\overline{ص} = ٦ \div ٩٠ = ١٥$$

$$٢,٧٥٧ = \sqrt{\frac{٣٨}{٥}} = ع$$

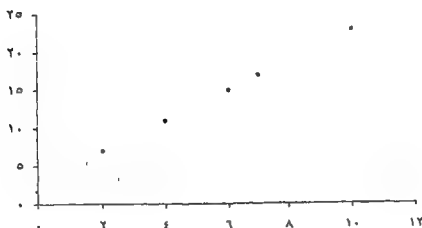
$$٥,٥١٤ = \sqrt{\frac{١٥٢}{٥}} = ع$$

بقسمة قيم العمود الرابع في الجدول السابق على الانحراف المعياري لقيم
س وقسمة قيم العمود الخامس على الانحراف المعياري لقيم ص نحصل
على الوحدات المعيارية للمتغيرين كما في الجدول التالي

مسلسل	س	ص	س-٦	ص-١٥	ز	ز
١	٢	٧	٤-	٨-	١,٦٤١-	١,٦٤١-
٢	١٠	٢٣	٤	٨	١,٦٤١	١,٦٤١
٣	٧	١٧	١	٢	٠,٣٦٢٧	٠,٣٦٢٧
٤	٤	١١	٢-	٤-	٠,٧٢٥٤-	٠,٧٢٥٤-
٥	٧	١٧	١	٢	٠,٣٦٢٧	٠,٣٦٢٧
٦	٦	١٥	٠	٠	٠	٠
المجموع	٣٦	٦٠	صفر	صفر	صفر	صفر

حيث أن جميع قيم الوحدات المعيارية للمتغيرين متساوية، فإن جميع مربعات الفروق في الصيغة (١-٧) سوف تكون مساوية للصفر وبالتالي تكون قيمة معامل بيرسمون للارتباط، $r = ١$.

ويرجع تساوي قيم الوحدات المعيارية للمتغيرين في هذا المثال إلى وجود علاقة خطية وطردية تامة بين المتغيرين مما يعني أن قيم أحد المتغيرين قد ضربت في مقدار ثابت وأضيف إليها ثابت آخر، وهذا كما سبق وأن رأينا لا يؤثر على القيم المعيارية، ويوضح شكل الانتشار التالي اتجاه العلاقة



نخلص من هذا إلى أنه إذا كانت هناك علاقة خطية وطريقة تامة بين المتغيرين، فإن قيم وحداتهما المعيارية سوف تكون دائما متساوية ويترتب على ذلك من المعادلة (٧_١) أن قيمة معامل الارتباط سوف تكون دائما مساوية للواحد الصحيح،

مثال ٧-٤

باستخدام المعادلة (٧_١) احسب معامل بيرسون للارتباط بين المتغيرين س و ص من البيانات الآتية

س	1	5	5	1
ص	14	16	16	14

الحل:

نترك للطالب على سبيل التدريب أن يقوم بحساب قيمة كل من الوسط الحسابي والانحراف المعياري لقيم المتغيرين ليجد أن،

$$\bar{S} = 4 \quad \bar{V} = 8 \quad \sigma_S = 3 \quad \sigma_V = 6$$

ونقوم في الجدول التالي بحساب القيم المعيارية ومربعات الفروق بينها تمهيدا لاستخدام المعادلة (٧_١) لحساب قيمة ر.

من	ص	Z	Z	من	ص
1	14	1-	1-	1	14
0	6	0,3333-	0,3333-	0,3333-	0,3333-
0	6	0,3333-	0,3333-	0,3333-	0,3333-
1	14	1-	1-	1	14
8	0	1,3334-	1,3334-	1,3334-	1,3334-
20	40	صفر	صفر	صفر	16

من معادلة (١-٧) نجد أن

$$r = 1 - \frac{16}{4 \times 2}$$

$$r = 1 - 2$$

$$r = -1$$

وتوضح هذه النتيجة أن هناك علاقة عكسية تامة بين المتغيرين، ويظهر من الجدول السابق أن هذه الحالة سوف تحدث عندما تتساوى القيم العددية للوحدات المعيارية وتختلف إشاراتهما، فإذا كانت $Z = -Z$ لجميع القيم فإنه يمكن كتابة المعادلة (١-٧) على الصورة

$$r = 1 - \frac{1}{(n-1)^2} \text{ مج } (Z - (Z^-))$$

$$r = 1 - \frac{4}{\sum (Z_i)^2}$$

وحيث أن:

$$1 = \frac{\sum Z_i^2}{(n-1)}$$

فإن

$$r = 1 - 2 = -1$$

من المثالين السابقين يمكن أيضا استنتاج أنه كلما اقتربت القيم المعيارية للمتغيرين من بعضها البعض كلما قلت الفروق بينها والتكرب مجموع المربعات في المعادلة (١-٧) من الصفر وبالتالي اقتربت قيمة معامل الارتباط من الواحد الصحيح، ومن ناحية أخرى كلما اقتربت القيم المطلقة للوحدات المعيارية وكانت إشاراتها مختلفة كلما اقتربت قيمة معامل الارتباط من سالب واحد صحيح ليندل على وجود علاقة عكسية قوية بين المتغيرين،

مثال ٧-٥

استخدم المعادلة (١-٧) لحساب معامل الارتباط بين المتغيرين س و ص من البيانات الآتية

س	3	5	4	8	10
ص	12	8	10	6	13

الحل:

عند حساب الوسط الحسابي والانحراف المعياري لكل من المتغيرين نجد أن

$$\bar{ن} = 6 \quad \bar{ص} = 9,8 \quad ع = 2,92 \quad ع = 2,86$$

وبالتالي تكون الوحدات المعيارية ومربعات الفروق بينها كما في الجدول التالي

س	ص	Z	Z _ص	Z _س - Z _ص	(Z _س - Z _ص) ²
3	15	1,02899-	0,76827	1,79726-	3,23016
5	8	0,34300-	0,62859-	0,28559	0,08156
4	10	0,68599-	0,69884	0,75584	0,57129
8	6	0,68599	1,32702-	2,01301	4,05222
10	13	1,37199	1,11749	0,25450-	0,06477
		صفر	صفر	صفر	8,000

من معادلة (٧-١) نجد أن

$$r = -1 = \frac{8 \times 1}{(4)^2} = 1 - 1 = \text{صفر}$$

أي أنه لا توجد علاقة خطية بين المتغيرين،

وعند النظر إلى قيم الوحدات المعيارية للمتغيرين نجد في هذه الحالة أن الإشارات تكون عكسية في بعض الأحيان وتكون متماثلة في أحيان أخرى، كذلك نجد أن بعض القيم المعيارية الكبيرة لأحد المتغيرين يناظرها قيم معيارية، أحيانا صغيرة وأحيانا أخرى كبيرة للمتغير الأخر، ويعني هذا أن معامل الارتباط سوف يكون مساويا للصفر أو قريبا منه إذا كان هناك نمط عشوائي لارتباط القيم المعيارية ببعضها البعض وأيضا إشاراتها.

نتناول فيما يلي بعض الأمثلة لتوضيح كيفية حساب قيمة معامل بيرسون للارتباط باستخدام المعاليتين (٥_٧) و(٨_٧)

مثال ٧-٦

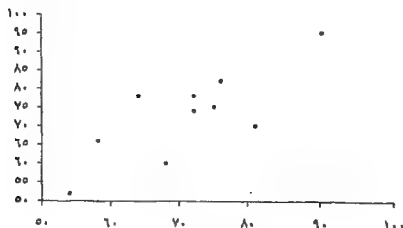
جمعت البيانات التالية عن درجات عينة من عشرة طلاب في امتحانين

للمحاسبة والإدارة

رقم الطالب	محاسبة	إدارة	رقم الطالب	محاسبة	إدارة
1	72	78	6	54	52
2	76	82	7	58	66
3	64	78	8	72	74
4	68	60	9	90	95
5	81	70	10	75	75

والمطلوب:

- 1- رسم شكل الانتشار بين المتغيرين والتعليق عليه من حيث درجة الارتباط الخطي واتجاه العلاقة،
 - 2- حساب قيمة معامل بيرسون للارتباط بين درجات الطلبة في امتحان المحاسبة و امتحان الإدارة،
- الحل:



1- يوضح شكل الانتشار وجود اتجاه علاقة موجبة بين درجات الطلبة في المادتين ولكننا لا نتوقع هنا أن تكون قوية جداً، ويتضح ذلك من تباعد النقاط إلى حد ما، والأساس المنطقي الذي يمكن الاستناد إليه هو أن الطالب المتفوق يكون متفوقاً في جميع الحالات والطلاب محدود المستوى يكون أداءه على مستوى ضعيف في جميع المواد، ولكن من جهة أخرى قد يؤدي ميل بعض الطلبة نحو العلوم الإدارية إلى التفوق فيها والحصول على درجات مرتفعة في امتحاناتها أكثر مما يحصلون عليه في المواد الأخرى، ويحدث العكس في حالة ميل الطالب نحو العلوم المحاسبية، وتتسبب مثل هذه الميول المختلفة في انحراف القيم المشاهدة للدرجات عن الأساس المنطقي وبالتالي تضعف من قوة العلاقة بين الدرجات في المادتين .

لحساب معامل الارتباط باستخدام المعادلة (٧_٨) نعتبر أن س تمثل درجات الطلبة في المحاسبة وص تمثل درجاتهم في الإدارة، بعد ذلك نقوم بضرب كل قيمة من قيم س في قيمة ص المناظرة لها، وكذلك نقوم بإيجاد مربع كل قيمة من قيم س ومربع كل قيمة من قيم ص ثم نأخذ مجاميع الأعمدة المختلفة ونعوض بها في صيغة المعادلة،

رقم الطالب	م	ص	م ص	م ص	ص
١	٧٢	٧٨	٥٦١٦	٥١٨٤	٦٠٨٤
٢	٧٦	٨٢	٦٢٣٢	٥٧٧٦	٦٧٢٤
٣	٦٤	٧٨	٤٩٩٢	٤٠٩٦	٦٠٨٤
٤	٦٨	٦٠	٤٠٨٠	٤٦٢٤	٣٦٠٠
٥	٨١	٧٠	٥٦٧٠	٦٥٦١	٤٩٠٠
٦	٥٤	٥٢	٢٨٠٨	٢٩١٦	٢٧٠٤
٧	٥٨	٦٦	٣٨٢٨	٣٣٦٤	٤٣٥٦
٨	٧٢	٧٤	٥٣٢٨	٥١٨٤	٥٤٧٦
٩	٩٠	٩٥	٨٥٥٠	٨١٠٠	٩٠٢٥
١٠	٧٥	٧٥	٥٦٢٥	٥٦٢٥	٥٦٢٥
مجموع	٧١٠	٧٣٠	٥٢٧٢٩	٥١٤٣٠	٥٤٥٧٨

$$\left[\frac{\text{مجموع م} - (\text{مجموع ص})}{\text{ن}} \right] \frac{1}{1-\alpha}$$

=

$$\sqrt{\frac{1}{\text{ن}} \left[\text{مجموع م} - (\text{مجموع ص}) \right] \frac{1}{1-\alpha}} \quad \sqrt{\frac{1}{\text{ن}} \left[\text{مجموع م} - (\text{مجموع ص}) \right] \frac{1}{1-\alpha}}$$

$$r = \frac{\left[\frac{(730)(710)}{10} - 52729 \right] \frac{1}{9}}{\sqrt{\left[\frac{(730)^2}{10} - 54078 \right] \frac{1}{9}} \sqrt{\left[\frac{(710)^2}{10} - 51430 \right] \frac{1}{9}}}$$

$$r = \frac{99,889}{11,963 \times 10,640} = 0,7840 =$$

وتؤكد قيمة معامل بيرسون للارتباط النتيجة التي توقعناها من قبل وهي وجود علاقة طردية بين درجات الطلبة في المادتين ولكنها ليست عالية القوة.

يوضح المثال السابق حجم العمليات الحسابية التي نجريها أثناء إيجاد حواصل الضرب ومربعات الأرقام، وحتى إذا كنا نستخدم الآلات الحاسبة فإن تسجيل العديد من الأرقام يزيد من احتمالات الخطأ، كذلك يلاحظ أن الوسط الحسابي لكل من المتغيرين يكون عددا صحيحا وبالتالي فإن استخدام المعادلة (٧-٥) في إيجاد قيمة معامل الارتباط تعتبر أفضل كثيرا في هذه الحالة كما يتضح مما يلي،
(عادة حل المثال باستخدام المعادلة (٧-٥))
في هذه الحالة تكون الخطوات المتبعة في الحل كما يلي:

- 1- نقوم بحساب الوسيط الحسابي للمتغيرين،
- 2- نوجد انحرافات قيم \bar{S} عن \bar{S} ،
- 3- نوجد انحرافات قيم \bar{V} عن \bar{V} ،
- 4- نوجد حواصل ضرب الانحرافات المتناظرة،
- 5- نقوم بإيجاد مربعات انحرافات قيم \bar{S} وقيم \bar{V} ،
- 6- نأخذ المجاميع المختلفة ونعوض في العلاقة

$$\bar{V} = 73$$

$$\bar{S} = 71$$

س	ص	س - \bar{S}	ص - \bar{V}	(س - \bar{S}) ²	(ص - \bar{V}) ²	(س - \bar{S}) × (ص - \bar{V})
٧٢	٧٨	١	٥	١	٢٥	٥
٧٦	٨٢	٥	٩	٢٥	٨١	٤٥
٦٤	٧٨	-٧	٥	٤٩	٢٥	-٣٥
٦٨	٦٠	-٣	-١٣	٩	١٦٩	٣٩
٨١	٧٠	١٠	-٣	١٠٠	٩	-٣٠
٥٤	٥٢	-١٧	-٢١	٢٨٩	٤٤١	٣٥٧
٥٨	٦٦	-١٣	-٧	١٦٩	٤٩	٩١
٧٢	٧٤	١	١	١	١	١
٩٠	٩٥	١٩	٢٢	٣٦١	٤٨٤	٤١٨
٧٥	٧٥	٤	٢	١٦	٤	٨
مجموع	١٢٨٨	١٠٢٠	٨٩٩	٠	٠	٠

$$r = \frac{\frac{1}{n-1} \text{مجم}(س-س)(ص-ص)}{\sqrt{\frac{\text{مجم}(س-س)^2}{n-1}} \sqrt{\frac{\text{مجم}(ص-ص)^2}{n-1}}}$$

$$r = \frac{899}{1488 \sqrt{1.40} \sqrt{1.684}} = 0.684$$

يلاحظ أن كلا من البسط والمقام في صيغة الارتباط دائماً ما يحتويان على نفس العامل (ن-١) وبالتالي يمكن إهماله عند الحساب كما فعلنا في الخطوة السابقة،

مثال ٧-٧

في دراسة للعلاقة بين طول لاعب كرة السلة وعدد افرميات الثلاثة التي يحرزها اللاعب جمعت البيانات التالية من عينة تتكون من ثمانية لاعبين حيث سجلت أطوالهم (س) وعدد الرميات الثلاثة التي أحرزها كل منهم في آخر أربع مباريات (ص)،

والمطلوب حساب معامل بيرسون للارتباط بين المتغيرين،

س	178	184	180	189	176	192	187	180
ص	5	7	6	9	4	10	8	9

الحل:

نقوم أولاً بحل المثال باستخدام القيم الأصلية للملاحظات، وبملاحظة أن قيم المتغير s كبيرة وأن وسطه الحسابي هو رقم كسري (١٨٣,٧٥) ، فإننا سوف نقوم بإعادة الحل بعد طرح مقدار ثابت من جميع قسَم s وليكن (١٨٠) ، وحيث أن قيم s صغيرة ولا توجد صعوبة في التعامل معها حسابياً فإننا نتركها كما هي،

s	s	s	s	s
١٧٨	٥	٨٩٠	٣١٦٨٤	٢٥
١٨٤	٧	١٢٨٨	٣٣٨٥٦	٤٩
١٨٠	٦	١٠٨٠	٣٢٤٠٠	٣٦
١٨٩	٩	١٧٠١	٣٥٧٢١	٨١
١٧٦	٤	٧٠٤	٣٠٩٧٦	١٦
١٩٢	١٠	١٩٢٠	٣٦٨٦٤	١٠٠
١٨٧	٨	١٤٩٦	٣٤٩٦٩	٦٤
١٨٠	٩	١٦٢٠	٣٢٤٠٠	٨١
١٤٦٦	٥٨	١٠٦٩٩	٢٦٨٨٧٠	٤٥٢

$$ع\ س\ س = \left[\frac{٥٨ \times ١٤٦٦}{٨} - ١٠٦٩٩ \right] \frac{١}{٧} = ١٠٠٠٧١٤ =$$

$$\left(\frac{2(1466)}{8} - 268870 \right) \frac{1}{\sqrt{}} = عس$$

$$= 0,676$$

$$\left(\frac{2(58)}{8} - 452 \right) \frac{1}{\sqrt{}} = عس$$

$$= 2,121$$

وحيث أن

$$\frac{عس}{عس} = ر$$

فإن

$$\frac{10,0714}{2,121 \times 0,676} = ر$$

$$= 0,837$$

يلاحظ في هذا المثال أننا قد قمنا بحساب مكونات صيغة معامل الارتباط كل على حدة ثم استخدمنا المعادلة (٦_٧) لحساب قيمته،

نقوم الآن بحل نفس المثال بطرح القيمة ١٨٠ كوسط فرضي من قيم س

$$س - ١٨٠$$

ثم نجري نفس العمليات السابقة باستخدام عمودي س و س لحساب قيم معامل الارتباط،

س	ح	ص	س ص	ح ص	ص
١٧٨	٢-	٥	١٠-	٤	٢٥
١٨٤	٤	٧	٢٨	١٦	٤٩
١٨٠	٠	٦	٠	٠	٣٦
١٨٩	٩	٩	٨١	٨١	٨١
١٧٦	٤-	٤	١٦-	١٦	١٦
١٩٢	١٢	١٠	١٢٠	١٤٤	١٠٠
١٨٧	٧	٨	٥٦	٤٩	٦٤
١٨٠	٠	٩	٠	٠	٨١
مج	٢٦	٥٨	٢٥٩	٣١٠	٤٥٢

$$r = \frac{\left[\frac{(58)(26)}{8} - 259 \right] \frac{1}{7}}{\sqrt{\left[\frac{(58)^2}{8} - 452 \right] \frac{1}{7}} \sqrt{\left[\frac{(26)^2}{8} - 310 \right] \frac{1}{7}}}$$

$$r = \frac{10,071}{2,121 \times 0,676} = 8,37$$

من هذا يرى القارئ، وكما ذكرنا من قبل، أن قيمة معامل بيرسون للارتباط لا تتأثر عند طرح أو جمع مقدار ثابت على قيم أحد المتغيرين أو كلاهما، وتوضح قيمة معامل الارتباط وجود علاقة طردية قوية بين طول اللاعب وعدد الرميات الثلاثية التي ينجح في تصويبها،

سبق أن نذكر أن معامل بيرسون للارتباط يقيس قوة العلاقة الخطية بين متغيرين وبالتالي يمكن أن نجد أن قيمة معامل بيرسون للارتباط تساوي صفر أو تكون قريبة منه ومع ذلك تكون هناك علاقة شبه تامة بين المتغيرين ولكنها غير خطية كما يتضح من المثال التالي.

مثال ٧-٨

احسب معامل بيرسون للارتباط بين المتغيرين س و ص باستخدام البيانات التالية:

س	1	2	3	5	6	7
ص	11	6	3	3	6	11

الحل:

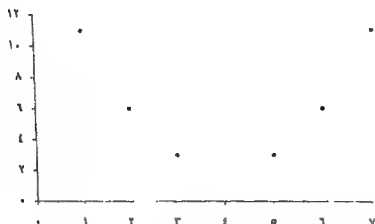
نقوم بحساب القيم المختلفة في الجدول التالي

س	ص	س ص	س ^٢	ص ^٢
١	١١	١١	١	١٢١
٢	٦	١٢	٤	٣٦
٣	٣	٩	٩	٩
٥	٣	١٥	٢٥	٩
٦	٦	٣٦	٣٦	٣٦
٧	١١	٧٧	٤٩	١٢١
٢٤	٤٠	١٦٠	١٢٤	٣٣٢

$$r = \frac{\frac{(40)(24) - 160}{6}}{\sqrt{\frac{(40) - 332}{6}} \sqrt{\frac{(24) - 124}{6}}}$$

$r =$ صفر.

من ناحية أخرى، إذا نظرنا إلى شكل الانتشار نجده على صورة منحنى معادلة من الدرجة الثانية كما في الشكل التالي



في المثال التالي نوضح تأثير القيم المتطرفة على قيمة معامل بيرسون للارتباط وذلك بحساب قيمة المعامل من البيانات الكاملة ثم حساب قيمته بعد حذفها،

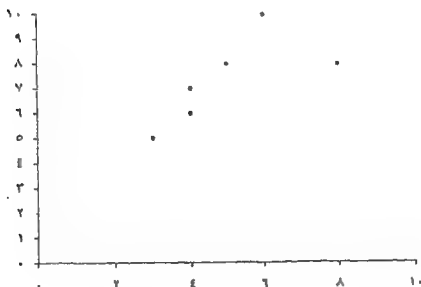
مثال ٧-٩

ارسم شكل الانتشار ثم احسب قيمة معامل بيرسون للارتباط باستخدام البيانات التالية

س	3	4	4	4	5	6	8
ص	5	6	7	6	8	10	8

الحل :

نلاحظ من شكل الانتشار التالي أن جميع النقاط تقع إلى حد كبير في اتجاه خط مستقيم ما عدا الملاحظة الأخيرة حيث يكون مستواها منخفض بصورة واضحة عن اتجاه هذا الخط، بالتالي تعتبر الملاحظة الأخيرة قيمة شاذة عن الاتجاه الذي تأخذه باقي القيم،



س	ص	س ص	س	ص
٣	٥	١٥	٩	٢٥
٤	٦	٢٤	١٦	٣٦
٤	٧	٢٨	١٦	٤٩
٤	٦	٢٤	١٦	٣٦
٥	٨	٤٠	٢٥	٦٤
٦	١٠	٦٠	٣٦	١٠٠
٨	٨	٦٤	٦٤	٦٤
٣٤	٥٠	٢٥٥	١٨٢	٣٧٤

تكون قيمة معامل الارتباط في وجود القيمة المتطرفة

$$r = \frac{\frac{(50 \times 34) - 2500}{\sqrt{\frac{2(50) - 374}{\sqrt{2(34) - 182}}}}}{\sqrt{\frac{2(50) - 374}{\sqrt{2(34) - 182}}}}} = 0.7203$$

وإذا قمنا بحذف المشاهدة الأخيرة ثم أعدنا تكوين جدول البيانات

س	ص	س ص	س	ص
٣	٥	١٥	٩	٢٥
٤	٦	٢٤	١٦	٣٦
٤	٧	٢٨	١٦	٤٩
٤	٦	٢٤	١٦	٣٦
٥	٨	٤٠	٢٥	٦٤
٦	١٠	٦٠	٣٦	١٠٠
٢٦	٤٢	١٩١	١١٨	٣١٠

سوف نجد أن قيمة معامل الارتباط

$$r = 0,974$$

يوضح هذا المثال التأثير الشديد لوجود مشاهدة شاذة في البيانات على القيمة المحسوبة لمعامل الارتباط حيث زادت قيمته بعد استبعاد هذه المشاهدة من 0,72 إلى 0,974

٧-٤ معامل التحديد Coefficient of Determination

نكرنا من قبل أن أحد أهداف تحليل العلاقات بين المتغيرات هو التعرف على دور المتغير أو المتغيرات المستقلة في تحديد قيم المتغير التابع وتفسير الاختلافات المشاهدة في قيمه، وعند دراسة العلاقة بين متغيرين فقط لا غير، فإبنا نطلق على مربع قيمة معامل بيرسون للارتباط اسم معامل التحديد (r^2)، ويقاس معامل التحديد نسبة الاختلافات المشاهدة في قيم المتغير التابع والتي يمكن تفسيرها من خلال تأثير المتغير المستقل عليه، فعلى سبيل المثال، إذا كانت قيمة معامل الارتباط بين المتغير التابع ص، والمتغير المستقل س، هي 0,9، فإن قيمة معامل التحديد تكون 0,81، وتعني هذه القيمة أن 81% من الاختلافات المشاهدة في قيم ص تفسر باختلافات القيم المناظرة للمتغير المستقل س، وتمثل النسبة المتممة، 19%، تأثير بعض المتغيرات الأخرى التي تربطها علاقة بالمتغير التابع ص والتي لم تدخل في نموذج الدراسة، وفي حالة تحليل العلاقة بين متغير تابع وأكثر من متغير مستقل، فإبنا نستخدم تعبير معامل التحديد المتعدد وسوف نشرح فيما بعد الأسلوب العام لحساب قيم معاملات التحديد بصرف النظر عن عدد المتغيرات المستقلة في النموذج، ونرمز لمعامل التحديد بين

المتغير التابع ص ومتغيرين مستقلين، س و ع على سبيل المثال، بوضع رموز المتغيرات المستقلة كأدلة سفلية بين قوسين ر^٢ ص (س ع)، ويطلق على الجذر التربيعي لمعامل التحديد المتعدد اسم معامل الارتباط المتعدد والذي يقيس قوة العلاقة الخطية بين المتغير التابع ومجموعة المتغيرات المستقلة في النموذج، وفي الحالة الخاصة بدراسة العلاقة الخطية بين متغير تابع ص ومتغيرين مستقلين س و ع، يمكن حساب قيمة معامل التحديد المتعدد باستخدام معاملات بيرسون للارتباط بين كل متغيرين باستخدام العلاقة

معامل التحديد المتعدد =

$$\frac{r_{ص س} + r_{ص ع}^2 - 2r_{ص س} r_{ص ع} r_{س ع}}{1 - r_{س ع}^2}$$

معامل الارتباط المتعدد = $\sqrt{\text{معامل التحديد المتعدد}}$

مثال ٧-١٠

في نموذج لتحليل العلاقة الخطية بين متغير تابع ص ومتغيرين مستقلين س و ع كانت معاملات الارتباط بين كل زوج من المتغيرات كما يلي:

$$r_{ص س} = 0,747 \quad r_{ص ع} = 0,648 \quad r_{س ع} = 0,764$$

احسب قيمة كل من معامل التحديد المتعدد ومعامل الارتباط المتعدد،

الحل:

باستخدام الصيغة السابقة تكون قيمة معامل التحديد المتعدد

$$\text{لـ س} = ٠,٧٤٧ \quad \text{لـ ع} = ٠,٦٤٨ \quad \text{لـ س ع} = ٠,٧٦٤$$

$$R^2 = \frac{٠,٧٦٤ \times ٠,٦٤٨ \times ٢ - (٠,٦٤٨)^2 - (٠,٧٤٧)^2}{(٠,٧٦٤)^2 - ١} = ٠,٥٧٢$$

$$\text{لـ س ع} = ٠,٧٥٧ = \sqrt{٠,٥٧٢}$$

تعني قيمة معامل الارتباط الكلي أنه يوجد اتجاه علاقة خطية بين المتغيرين ولكنها متوسطة وليست قوية، وإذا نظرنا إلى معامل التحديد المتعدد نجد أن المتغيرين المستقلين س و ع يفسران فقط ٥٧,٢% من تغيرات المتغير التابع ص، وإذا ما حسبنا نسبة تفسير كل متغير مستقل لتغيرات المتغير التابع على حده نجد أن $R^2_{\text{لـ س}} = ٥٥,٨\%$ وأن $R^2_{\text{لـ ع}} = ٤٢\%$ ، نستنتج من هذا أن نمية تفسير المتغيرين المستقلين لتغيرات المتغير التابع ليست بصفة عامة محصلة مجموع نسبة تفسير كل متغير مستقل على حده، ولن يحدث هذا إلا إذا كانت قيمة معامل الارتباط بين المتغيرين المستقلين مساوية للصفر كما يتضح من الصيغة السابقة، ويمكن تفسير العبارة السابقة بأنه إذا لم تكن قيمة معامل الارتباط بين المتغيرين المستقلين مساوية للصفر، فإن هذا يتضمن وجود اتجاه مشترك لهما للتأثير على

المتغير التابع بنفس النمط، ويظهر هذا الاتجاه المشترك في قيمة معامل التحديد الخاصة بكل متغير مستقل مع المتغير التابع، وبالتالي نجد أن مجموع قيمتي معامل التحديد تكون أكبر من قيمة معامل التحديد الكلي، نقدم فيما يلي الأسلوب العام لحساب قيمة معامل التحديد المتعدد ونشرح كيفية النظر إلى قيمته كنسبة تفسير المتغيرات المستقلة لمتغيرات قيم المتغير التابع وذلك من خلال تحليل المثال التالي.

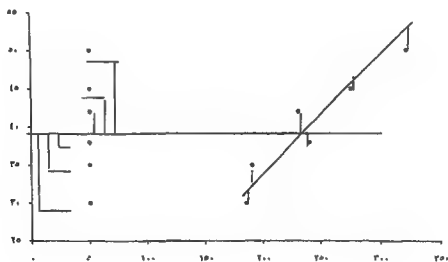
مثال ٧-١١

القائمة التالية لأسعار بيع ستة كتب مختلفة،

من هذه البيانات نجد أن متوسط سعر الكتاب يكون، $\bar{X} = 40$ ، وتكون

انحرافات أسعار الكتب عن وسطها الحسابي على الصورة

السعر (ص)	30	35	38	42	45	50
ص - 40	10-	5-	2-	2	5	10



تظهر في الصف الثالث من الجدول السابق بمثابة أخطاء عشوائية، ولما كان مجموع هذه الانحرافات يساوي الصفر، فإننا نستخدم مجموع مربعاتها كمقياس لحجم الاختلافات الكلية (total variation)، وسوف نطلق على مجموع مربعات انحرافات القيم عن وسطها الحسابي اسم مجمع المربعات الكلي (sum of squares of total variation)، أي أن

$$\text{مجموع المربعات الكلي} = \text{مج} - \frac{(\text{ص})^2}{\text{م}}$$

يلاحظ أن مجموع المربعات الكلي هو البسط في صيغة تباين ص، وتظهر أحجام الاختلافات الكلية في شكل الانتشار السابق من خلال الأبعاد الرأسية للنقاط الموقعة إلى يمين المحور الرأسي مباشرة عن الخط الأفقي الذي يمثل قيمة الوسط الحسابي (٤٠)، والآن إذا نظرنا إلى عدد صفحات الكتاب كأحد العوامل المؤثرة في سعر بيعه وكانت البيانات كما يلي

م	1	2	3	4	5	6
عدد الصفحات (م)	185	190	240	230	275	320
السعر (ص)	30	35	38	42	45	50

بالعودة إلى شكل الانتشار سنجد أن أننى نقطة في النقاط الرأسية سوف تنتقل يمينا لتوقع أعلى القيمة ١٨٥، كذلك سوف تنتقل القيمة التالية (٣٥) لتوقع أعلى القيمة ١٩٠ وهكذا، مع أخذ عدد صفحات الكتاب في الاعتبار،

يبين شكل الانتشار اتجاه علاقة خطية طردية نستطيع من خلالها القول بأن سعر الكتاب الأول (٣٠) جاء منخفضاً لأن عدد صفحاته هي الأقل، بينما سعر الكتاب السادس هو الأعلى لأن عدد صفحاته هي الأكبر، ولكون العلاقة غير تامة تظهر في الشكل انحرافات رأسية بين النقاط المتشابهة والنقاط المقابلة لها على خط العلاقة وهذه الانحرافات مرة أخرى تمثل أخطاء عشوائية لا يمكن تفسيرها في ظل غياب معطومات عن أية متغيرات مستقلة أخرى لها تأثير على سعر بيع الكتاب وتطلق عليها اسم التغيرات غير المفسرة (unexplained variation)، وإذا رمزنا للتقسيم المقدرة للسعر على خط العلاقة بالرمز \hat{y} ، فإن التغيرات غير المفسرة والتي تستخدم كتقدير للأخطاء العشوائية تحسب بالفروق التغيرات غير المفسرة = $\hat{y} - y$ (ص - ص)

ويلاحظ أن أحجام الأخطاء في ظل النظر إلى أسعار الكتب من خلال أعداد صفحاتها أصبحت أقل كثيراً مما كانت عليه عند النظر إلى الأسعار فقط، وتمثل الفروق بين التغيرات الكلية والتغيرات غير المفسرة بعد أخذ عدد صفحات الكتاب في الاعتبار، مساهمة قيم المتغير المستقل في تفسير اختلافات قيم المتغير التابع ونطلق عليها اسم التغيرات المفسرة (explained variation)، أي أن

التغيرات المفسرة = التغيرات الكلية - التغيرات غير المفسرة

$$= (\text{ص} - \text{ص}) - (\text{ص} - \text{ص})$$

$$= (\text{ص} - \text{ص})$$

وإذا ما قمنا بحساب مجموع التغيرات المفسرة ومجموع التغيرات غير المفسرة فسوف نجد أن كل منهما يساوي الصفر، بالتالي نقوم بأخذ مجموع مربعات نوعي التغيرات السابقين لنحصل على

$$\text{مجموع المربعات المفسر} = \text{مج} (ص - \text{ص}^2)$$

$$\text{مجموع المربعات غير المفسر} = \text{مج} (ص - \text{ص}^2)$$

وسوف نشاهد دائما صحة العلاقة التالية:

$$\text{مجموع المربعات الكلي} = \text{مجموع المربعات المفسر} + \text{مجموع المربعات غير المفسرة}$$

$$\text{مج} (ص - \text{ص}^2) = \text{مج} (ص - \text{ص}^2) + \text{مج} (ص - \text{ص}^2)$$

والنسبة بين مجموع المربعات المفسر إلى مجموع المربعات الكلي تمثل نسبة مساهمة المتغير المستقل في تفسير تغيرات قيم المتغير التابع وبالتالي فإنها تعطي قيمة معامل التحديد، أي أنه يمكن حساب قيمة معامل التحديد باستخدام الصيغة

$$\text{معامل التحديد} = r^2 = \frac{\text{مجموع المربعات المفسر}}{\text{مجموع المربعات الكلي}}$$

$$\text{مج} (ص - \text{ص}^2)$$

$$\text{معامل التحديد} = r^2 = \frac{\text{مج} (ص - \text{ص}^2)}{\text{مج} (ص - \text{ص}^2)}$$

ينضح من الصيغ السابقة ضرورة معرفة معادلة الخط المستقيم الذي يمثل العلاقة بين المتغير التابع والمتغيرات المستقلة وذلك حتى نتمكن من

التعويض في هذه المعادلة بقيم المتغير أو المتغيرات المستقلة لحساب القيم المقدرة، وسوف نتناول في الفصل القادم كيفية تقدير مثل هذه المعادلات، دعنا نفترض الآن أن العلاقة المقدرة بين أسعار الكتب (ص) وعدد صفحاتها (س) تكون على الصورة:

$$\text{ص} = ٨,٢١١٣ + ٠,١٣٥٥ س$$

وبحساب معامل بيرسون للارتباط بين المتغيرين نجد أن قيمته هي 0,971 وبالتالي تكون قيمة معامل التحديد 0,943 ، ونقوم الآن بتطبيق الصيغة العامة لحساب معامل التحديد كنسبة بين مجاميع المربعات للنرى - أننا سوف نصل إلى نفس النتيجة،
أول خطوة نتبعها هي التعويض بالقيم المشاهدة للمتغير المستقل س في المعادلة السابقة للحصول على القيم المقدرة ص،

س	ص = ٨,٢١١ + ٠,١٣٢٥ س	ص
١٨٥	$= ١٨٥ \times ٠,١٣٢٥ + ٨,٢١١$	٣٢,٧١٥
١٩٠	$= ١٩٠ \times ٠,١٣٢٥ + ٨,٢١١$	٣٣,٣٧٧
٢٣٠	$= ٢٣٠ \times ٠,١٣٢٥ + ٨,٢١١$	٤٥,٠٠٠
٢٤٠	$= ٢٤٠ \times ٠,١٣٢٥ + ٨,٢١١$	٣٨,٦٧٦
٢٧٥	$= ٢٧٥ \times ٠,١٣٢٥ + ٨,٢١١$	٤٤,٦٣٦
٣٢٠	$= ٣٢٠ \times ٠,١٣٢٥ + ٨,٢١١$	٥٠,٥٩٦

ويظهر الجدول التالي التغيرات المختلفة ومربعاتها ومنه نلاحظ ما يلي:

1- مجاميع التغيرات الكلية، في العمود الثالث، والمفسرة بالعلاقة بين المتغيرين، في العمود الرابع، وغير المفسرة أو العشوائية، في العمود الخامس، كلها تساوي الصفر،

2- حاصل جمع مجموع مربعات التغيرات المفسرة ومجموع مربعات التغيرات غير المفسرة يساوي مجموع مربعات التغيرات الكلية،

$$258 = 14,828 + 243,172$$

3- معامل التحديد هو النسبة بين مجموع المربعات المفسر إلى مجموع المربعات الكلية

$$\text{معامل التحديد} = \frac{243,172}{258} = 0,943$$

وهي نفس القيمة السابقة،

4 - إن ارتفاع قيمة معامل التحديد بهذه الصورة وبالتالي انخفاض نسبة التغيرات غير المفسرة (5,7%) تعني أن النموذج الذي يربط بين سعر الكتاب وعدد صفحاته يضمن إعطاء تنبؤات عالية الدقة لتقدير سعر الكتاب إذا علمنا عدد صفحاته،

2 (ص-مق)	2 (مق-ص)	2 (ص-مق)	ص-مق	مق-ص	ص-مق	مق	ص
٦,٤٩٨	٥٥,٥١٧	١٠٠	٧,٥٤٩-	٧,٤٥١-	١٠-	٧٧,٥٤٩	٨٠
٣,١٤٦	٤٥,٨٨١	٢٥	١,٧٧٤	٦,٧٧٤-	٥-	٣٣,٢٢٦	٢٥
١,٤١٦	١,٨٣٥	٤	١,٦٤٥-	١,٣٥٥-	٢-	٣٨,٦٢٥	٣٨
٤	٤	٤	٧	٠	٢	٤٠	٤٢
١,٠٦٦	٢٢,٤٨٢	٢٥	١,٢٥٨	٤,٧٤٧	٥	٤٤,٧٤٢	٤٥
١,٧٠٢	١١٧,٤٥٧	١٠٠	١,٨٣٨-	١٠,٨٣٨	١٠	٥٠,٨٣٨	٥٠
١٤,٨٢٨	٢٤٣,١٧٢	٢٥٨	صفر	صفر	صفر	٢٤٠	٢٤٠

The Spearman Ranks Correlation Coefficient

ذكرنا أن معامل بيرسون للارتباط الخطي بين متغيرين يتطلب أن يكونا كميين. وإذا كان أحد المتغيرين أو كلاهما نوعي ترتيبي فإننا نقوم في هذه الحالة باستخدام معامل آخر للارتباط يطلق عليه معامل سبيرمان للارتباط أو معامل ارتباط الرتب. كذلك ينصح أيضاً باستخدام معامل ارتباط الرتب في حالة البيانات الكمية التي تحتوي على بعض القيم الشاذة حيث يتأثر معامل بيرسون للارتباط بوجود مثل هذه القيم بينما يكون معامل ارتباط الرتب أقل تأثراً بها.

ولحساب معامل ارتباط الرتب نتبع ما يلي:

- 1- نعين رتبة لكل صفة أو لكل قيمة من قيم المتغير الأول وفقاً لترتيب تصاعدي (أو تنازلي) للصفات أو القيم.
- 2- نتبع نفس الأسلوب في تعيين رتب لصفات أو قيم المتغير الثاني.
- 3- نقوم بحساب الفروق المتناظرة بين رتب المتغيرين
- ف = رتبة المتغير الأول - الرتبة المناظرة للمتغير الثاني.
- 4- نقوم بإيجاد مربع كل فرق من الفروق السابقة ونأخذ المجموع (مجم ف^٢).
- 5- أخيراً نستخدم العلاقة التالية لحساب قيمة معامل سبيرمان لارتباط الرتب

$$r = 1 - \frac{6 \cdot \text{مجنف}^2}{n(n^2 - 1)}$$

يلاحظ أنه يمكن أيضاً الحصول على قيمة معامل سيرمان للارتباط دون أخذ الفروق ومربعاتها وذلك عن طريق حساب معامل بيرسون للارتباط بين رتب المتغيرين بشرط عدم وجود تكرار لقيم أو صفات كل من المتغيرين.

مثال ٧-١٤

جمعت البيانات التالية لدراسة العلاقة بين عدد سنوات الخبرة وتقدير الكفاءة لعينة من عمال مصنع معين.

عدد سنوات الخبرة	تقدير كفاءة العامل
٨	ممتاز
٥	جيد جداً
٧	جيد
٣	مقبول
١	ضعيف

والمطلوب حساب الارتباط بين مستوى الخبرة والكفاءة.

الحل:

حيث أن متغير الكفاءة يكون وصفي ترتيبى وليس كمي، فإننا نستخدم معامل سيرمان لقياس درجة الارتباط بين المتغيرين

سنوات الخبرة	تقدير كفاءة	رتب الخبرة	رتب الكفاءة	فروق الرتب	مربعات الفروق
٨	ممتاز	٥	٥	٠	٠
٥	جيد جدا	٣	٤	١-	١
٧	جيد	٤	٣	١	١
٣	ضعيف	٢	١	١	١
١	مقبول	١	٢	١-	١
مجـ				صفر	٤

يلاحظ في الجدول السابق:

- 1- عند تعيين رتب الخبرة أعطينا أقل القيم الرتبة الأولى ثم القيمة التالية لها في الصغر الرتبة الثانية وهكذا. وبالتالي لابد من اتباع نفس الأساس في تعيين رتب تقديرات الكفاءة حيث عينا الرتبة الأولى لأقل التقديرات والرتبة الثانية للتقدير التالي وهكذا.
- 2- إذا قمنا بتعيين الرتب بصورة صحيحة، فلا بد وأن نجد أن مجموع عمود الفروق يساوي صفر.

$$\text{معامل سيرمان للارتباط} = 1 - \frac{4 \times 6}{(1 - 25)5}$$

$$= 1 - 0,2 = 0,8$$

ويعني هذا أنه يوجد اتجاه علاقة طردية قوية بين عدد سنوات الخبرة وتقدير الكفاءة.

وحيث أنه لا توجد قيم مكررة في عدد سنوات الخبرة ولا توجد صفة مكررة في تقديرات الكفاءة، فإننا سوف نحصل على نفس النتيجة السابقة إذا ما قمنا بحساب معامل بيرسون للارتباط بين رتب المتغيرين. فإذا استخدمنا r_s للدلالة على رتب الخبرة و r_x للدلالة على رتب الكفاءة، نحصل على المعلومات التالية

س	ص	س ص	س ²	ص ²
٥	٥	٢٥	٢٥	٢٥
٣	٤	١٢	٩	١٦
٤	٣	١٢	١٦	٩
٢	١	٢	٤	١
١	٢	٢	١	٤
١٥	١٥	٥٣	٥٥	٥٥

معامل بيرسون للارتباط بين رتب المتغيرين

$$r_s = \frac{10 \times 10 - 53}{\sqrt{\frac{2(10) - 00}{0} \cdot \frac{2(10) - 00}{0}}} = \frac{47 - 53}{10} = -0,8$$

يلاحظ في معظم الأحيان عند حساب معامل ارتباط الرتب أن هناك بعض الصفات أو القيم يتكرر ظهورها أكثر من مرة وفي هذه الحالة نعين

لكل صفة أو قيمة مكررة متوسط الرتب التي تحتلها. فعلى سبيل المثال إذا أردنا تعيين رتب للقيم

٢	٤	٤	٥	٧	٧	٧	٨
---	---	---	---	---	---	---	---

فإن القيمة ٢ تأخذ الرتبة ١، القيمة ٤ تكرر ظهورها مرتين وتحتل الرتبتين الثانية والثالثة وبالتالي نقوم بإعطاء كل منهما الرتبة ٢,٥. بعد ذلك تأتي القيمة ٥ تحتل الرتبة الرابعة. وحيث أن القيمة ٧ تكرر ظهورها ثلاث مرات وتحتل الرتب ٥، ٦، ٧ فإننا نعين لكل منها متوسط الرتب الثلاث (٦). وأخيراً تحتل القيمة ٨ الرتبة الثامنة كما يتضح من الجدول التالي.

القيمة	٢	٤	٤	٥	٧	٧	٧	٨
الرتب	١	٢,٥	٢,٥	٤	٦	٦	٦	٨

مثال ٧-١٥

البيانات التالية تمثل تقديرات عينة من الطلبة في امتحانين للإحصاء والاقتصاد. والمطلوب حساب قيمة معامل الارتباط بين التقديرات في المادتين.

تقديرات الإحصاء	تقديرات الاقتصاد
جيد	جيد
جيد	مقبول
مقبول	جيد
مقبول	مقبول
ممتاز	جيد جدا
ضعيف	مقبول
ممتاز	ممتاز
جيد جدا	جيد

نقوم في هذا المثال بتعيين الرتب وفقا لترتيب تنازلي للتقديرات فنبدأ بإعطاء الرتب الأولى للتقدير ممتاز ثم الرتب التالية للتقدير جيد جدا وهكذا. يلاحظ أن التقدير ممتاز تكرر مرتين في تقديرات الإحصاء وبالتالي يحتل المرتبتين الأولى والثانية ويكون متوسط الرتب هو 1.5 . يأخذ الطالب الذي حصل على جيد جدا الرتبة الثالثة وهكذا. وتظهر رتب المادتين كما في الجدول التالي

تقديرات الإحصاء	تقديرات الاقتصاد	رتب الإحصاء	رتب الاقتصاد	ف	ف ^٢
جيد	جيد	٤	٤,٥	٠,٥	٠,٢٥
جيد	مقبول	٧	٤,٥	٢,٥-	٦,٢٥
مقبول	جيد	٤	٦,٥	٢,٥	٦,٢٥
مقبول	مقبول	٧	٦,٥	٠,٥-	٠,٢٥
ممتاز	جيد جدا	٢	١,٥	٠,٥-	٠,٢٥
ضعيف	مقبول	٨	٨	١	١
ممتاز	ممتاز	١	١,٥	٠,٥-	٠,٢٥
جيد جدا	جيد	٣	٣	١-	١
مجم				صفر	١٥,٥

معامل سبيرمان للارتباط

$$r = \frac{100 \times 6}{(1-64)^8} - 1$$

$$= 0,185 - 1 = -0,815$$

من هذا يتضح وجود علاقة طردية قوية بين المتغيرين. أي أن الطالب الذي يحصل على تقدير مرتفع في الإحصاء من المتوقع أن يكون تقديره مرتفع أيضا في الاقتصاد والعكس صحيح.

مثال ٧-١٩

نقوم في هذا المثال بحساب قيمة معامل ارتباط الرتب من بيانات مثال ٧-٩ والذي سبق أن ذكرنا أن المشاهدة الأخيرة في بياناته تعتبر قيمة غير

طبيعية وكان لاستبعاد هذه القيمة أثر كبير حيث ارتفعت قيمة معامل بيرسون للارتباط من ٠,٧٢ إلى ٠,٩٧ . ونعيد كتابة البيانات في الجدول التالي ثم نقوم بحساب قيمة معامل ارتباط الرتب دون استبعاد المشاهدة الأخيرة.

س	٣	٤	٤	٤	٥	٦	٨
ص	٥	٦	٧	٦	٨	١٠	٨

الحل

س	ص	رتب س	رتب ص	ف _١	ف _٢
٣	٥	١	١	٠	٠
٤	٦	٣	٢,٥	٠,٢٥	٠,٢٥
٤	٧	٣	٤	١	١
٤	٦	٣	٢,٥	٠,٢٥	٠,٢٥
٥	٨	٥	٥,٥	٠,٢٥	٠,٥
٦	١٠	٦	٧	١	١
٨	٨	٧	٥,٥	٢,٢٥	١,٥
مجـ				٥	صفر

$$r = \frac{5 \times 6}{(1 - 49)7} - 1 =$$

$$= 0,89 - 1 = -0,11$$

يتضح من هذا المثال أن قيمة معامل سبيرمان لارتباط الرتب لم تتأثر بدرجة كبيرة بوجود القيمة المتطرفة مثل معامل بيرسون للارتباط.

تمارين الفصل السابع

١- جمعت البيانات التالية لقياس العلاقة بين حجم الإنفاق الشهري على

السلع الغذائية والدخل الشهري لعينة من الأسر

440	390	480	720	420	680	330	540	الإنفاق
760	620	680	1200	640	900	560	800	الدخل

احسب معامل بيرسون للارتباط وعلق النتيجة.

٢- أعد حل التمرين السابق بعد طرح 330 كوسط فرضي من قيم الإنفاق

و 560 كوسط فرضي من قيم الدخل ثم قسمة جميع النواتج على 10.

٣- فيما يلي بيانات بتقديرات عينة من طلبة السنة الأولى في كلية التجارة

والنسب المئوية لدرجاتهم في امتحان الثانوية العامة

جيد	مقبول	ضعيف	جيد جدا	جيد	مقبول	جيد	التقدير
%84	%78	%80	%90	%84	%82	%87	النسبة

احسب مقياس ملائم للارتباط لقياس قوة العلاقة بين المتغيرين وعلق على

النتيجة.

الجداول

$$\Phi(z) = P(Z \leq z)$$

جدول التوزيع الطبيعي المعياري

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9966	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Selected Normal Percentiles

P	.90	.95	.975	.99	.995
Z	1.2816	1.6449	1.96	2.3263	2.5758

جدول توزیع

Degrees Of Freedom	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$T_{0.005}$
1	3.078	6.314	12.706	31.821	63.656
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
∞	1.282	1.645	1.960	2.326	2.576

جدول توزیع χ^2

Degrees Of Freedom	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	2.7055	3.8415	5.0239	6.6349	7.8794
2	4.6052	5.9915	7.3778	9.2104	10.5965
3	6.2514	7.8147	9.3484	11.3449	12.8381
4	7.7794	9.4877	11.1433	13.2767	14.8602
5	9.2363	11.0705	12.8325	15.0863	16.7496
6	10.6446	12.5916	14.4494	16.8119	18.5475
7	12.0170	14.0671	16.0128	18.4753	20.2777
8	13.3616	15.5073	17.5345	20.0902	21.9549
9	14.6837	16.9190	19.0228	21.6660	23.5893
10	15.9872	18.3070	20.4832	23.2093	25.1881
11	17.2750	19.6752	21.9200	24.7250	26.7569
12	18.5493	21.0261	23.3367	26.2170	28.2997
13	19.8119	22.3620	24.7356	27.6882	29.8193
14	21.0641	23.6848	26.1189	29.1412	31.3194
15	22.3071	24.9958	27.4884	30.5780	32.8015
16	23.5418	26.2962	28.8453	31.9999	34.2671
17	24.7690	27.5871	30.1910	33.4087	35.7184
18	25.9894	28.8693	31.5264	34.8052	37.1564
19	27.2036	30.1435	32.8523	36.1908	38.5821
20	28.4120	31.4104	34.1696	37.5663	39.9969
21	29.6151	32.6706	35.4789	38.9322	41.4009
22	30.8133	33.9245	36.7807	40.2894	42.7957
23	32.0069	35.1725	38.0756	41.6383	44.1814
24	33.1962	36.4150	39.3641	42.9798	45.5584
25	34.3816	37.6525	40.6465	44.3140	46.9280
26	35.5632	38.8851	41.9231	45.6416	48.2898
27	36.7412	40.1133	43.1945	46.9628	49.6450
28	37.9159	41.3372	44.4608	48.2782	50.9936
29	39.0875	42.5569	45.7223	49.5878	52.3355
30	40.2560	43.7730	46.9792	50.8922	53.6719
40	51.8050	55.7585	59.3417	63.6908	66.7660

$$\alpha = 0.05$$

v_1	Numerator Degrees of Freedom							
	9	10	11	12	15	20	24	30
v_2								
1	240.54	241.88	242.98	243.90	245.95	248.02	249.05	250.10
2	19.38	19.40	19.40	19.41	19.43	19.45	19.45	19.46
3	8.81	8.79	8.76	8.74	8.70	8.66	8.64	8.62
4	6.00	5.96	5.94	5.91	5.86	5.80	5.77	5.75
5	4.77	4.74	4.70	4.68	4.62	4.56	4.53	4.50
6	4.10	4.06	4.03	4.00	3.94	3.87	3.84	3.81
7	3.68	3.64	3.60	3.57	3.51	3.44	3.41	3.38
8	3.39	3.35	3.31	3.28	3.22	3.15	3.12	3.08
9	3.18	3.14	3.10	3.07	3.01	2.94	2.90	2.86
10	3.02	2.98	2.94	2.91	2.85	2.77	2.74	2.70
11	2.90	2.85	2.82	2.79	2.72	2.65	2.61	2.57
12	2.80	2.75	2.72	2.69	2.62	2.54	2.51	2.47
13	2.71	2.67	2.63	2.60	2.53	2.46	2.42	2.38
14	2.65	2.60	2.57	2.53	2.46	2.39	2.35	2.31
15	2.59	2.54	2.51	2.48	2.40	2.33	2.29	2.25
16	2.54	2.49	2.46	2.42	2.35	2.28	2.24	2.19
17	2.49	2.45	2.41	2.38	2.31	2.23	2.19	2.15
18	2.46	2.41	2.37	2.34	2.27	2.19	2.15	2.11
19	2.42	2.38	2.34	2.31	2.23	2.16	2.11	2.07
20	2.39	2.35	2.31	2.28	2.20	2.12	2.08	2.04
21	2.37	2.32	2.29	2.25	2.18	2.10	2.05	2.01
22	2.34	2.30	2.26	2.23	2.15	2.07	2.03	1.98
23	2.32	2.27	2.24	2.20	2.13	2.05	2.01	1.96
24	2.30	2.25	2.22	2.18	2.11	2.03	1.98	1.94
25	2.28	2.24	2.20	2.16	2.09	2.01	1.96	1.92
26	2.27	2.22	2.18	2.15	2.07	1.99	1.95	1.90
27	2.25	2.20	2.17	2.13	2.06	1.97	1.93	1.88
28	2.24	2.19	2.15	2.12	2.04	1.96	1.91	1.87
29	2.22	2.18	2.14	2.10	2.03	1.94	1.90	1.85
30	2.21	2.16	2.13	2.09	2.01	1.93	1.89	1.84

$$\alpha = 0.05$$

v_1	Numerator Degrees of Freedom							
	1	2	3	4	5	6	7	8
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27

$$\alpha = 0.01$$

v_1	Numerator Degrees of Freedom							
	9	10	11	12	15	20	24	30
1	6022.40	6055.93	6083.40	6106.68	6156.97	6208.66	6234.27	6260.35
2	99.39	99.40	99.41	99.42	99.43	99.45	99.46	99.47
3	27.34	27.23	27.13	27.05	26.87	26.69	26.60	26.50
4	14.66	14.55	14.45	14.37	14.20	14.02	13.93	13.84
5	10.16	10.05	9.96	9.89	9.72	9.55	9.47	9.38
6	7.98	7.87	7.79	7.72	7.56	7.40	7.31	7.23
7	6.72	6.62	6.54	6.47	6.31	6.16	6.07	5.99
8	5.91	5.81	5.73	5.67	5.52	5.36	5.28	5.20
9	5.35	5.26	5.18	5.11	4.96	4.81	4.73	4.65
10	4.94	4.85	4.77	4.71	4.56	4.41	4.33	4.25
11	4.63	4.54	4.46	4.40	4.25	4.10	4.02	3.94
12	4.39	4.30	4.22	4.16	4.01	3.86	3.78	3.70
13	4.19	4.10	4.02	3.96	3.82	3.66	3.59	3.51
14	4.03	3.94	3.86	3.80	3.66	3.51	3.43	3.35
15	3.89	3.80	3.73	3.67	3.52	3.37	3.29	3.21
16	3.78	3.69	3.62	3.55	3.41	3.26	3.18	3.10
17	3.68	3.59	3.52	3.46	3.31	3.16	3.08	3.00
18	3.60	3.51	3.43	3.37	3.23	3.08	3.00	2.92
19	3.52	3.43	3.36	3.30	3.15	3.00	2.92	2.84
20	3.46	3.37	3.29	3.23	3.09	2.94	2.86	2.78
21	3.40	3.31	3.2	3.17	3.03	2.88	2.80	2.72
22	3.35	3.26	3.18	3.12	2.98	2.83	2.75	2.67
23	3.30	3.21	3.14	3.07	2.93	2.78	2.70	2.62
24	3.26	3.17	3.09	3.03	2.89	2.74	2.66	2.58
25	3.22	3.13	3.06	2.99	2.85	2.70	2.62	2.54
26	3.18	3.09	3.02	2.96	2.81	2.66	2.58	2.50
27	3.15	3.06	2.99	2.93	2.78	2.63	2.55	2.47
28	3.12	3.03	2.96	2.90	2.75	2.60	2.52	2.44
29	3.09	3.00	2.93	2.87	2.73	2.57	2.49	2.41
30	3.07	2.98	2.91	2.84	2.70	2.55	2.47	2.39

$$\alpha = 0.01$$

v_1	Numerator Degrees of Freedom							
	1	2	3	4	5	6	7	8
1	4052.18	4999.34	5403.53	5624.26	5763.96	5858.95	5928.33	5980.95
2	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49
4	21.20	18.00	16.69	15.98	15.32	15.21	14.98	14.80
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17

المحتويات

المقدمة :	٣
الفصل الأول : توزيعات المعاينة	٥
الفصل الثاني : تقدير معالم المجتمع	٥٣
الفصل الثالث : اختبارات الفروض الإحصائية	٩٥
الفصل الرابع : أساليب الاستدلال الإحصائي للمقارنة بين معالم مجتمعين	١٤٧
الفصل الخامس : تحليل التباين	٢١١
الفصل السادس : الاستدلال الإحصائي باستخدام أسلوب كا ^٢	٢٤٥
الفصل السابع : الارتباط الخطي بين الظواهر	٢٩٥
الجداول :	٣٤٣
المحتويات :	٣٥٢



الناشر
مكتبة الوفاء القانونية
٠٠٢٠١٠ ٣٧٣٨٨٢٢
تليفاكس
٠٠٢٠٣/٥٤٠٤٤٨٠
الإسكندرية